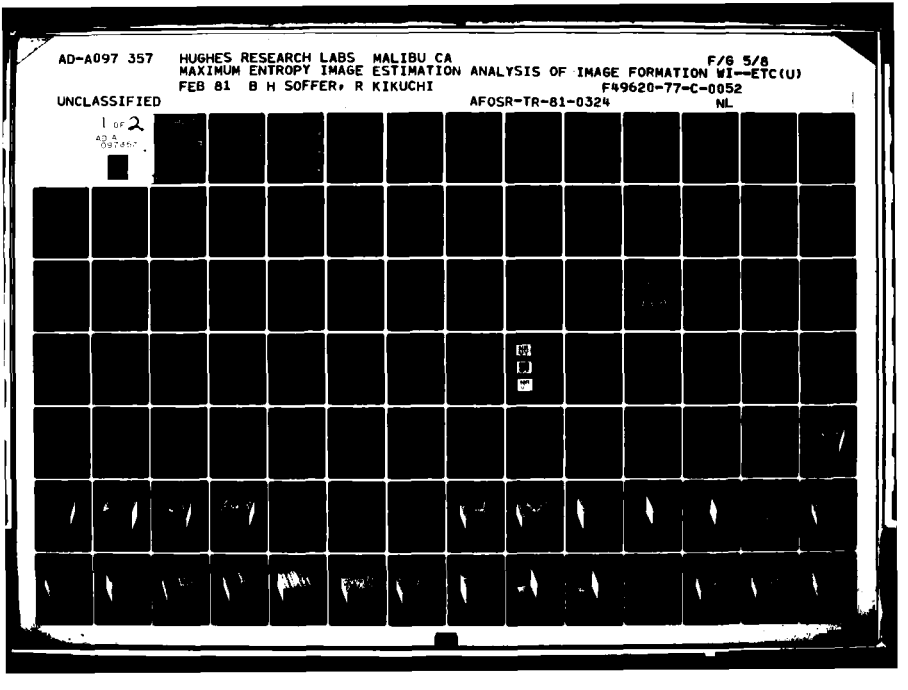
AD-A097 357 HUGHES RESEARCH LABS MALIBU CA F/G 5/8
MAXIMUM ENTROPY IMAGE ESTIMATION ANALYSIS OF IMAGE FORMATION WI--ETC(U)
FEB 81 B H SOFFER, R KIKUCHI F49620-77-C-0052
UNCLASSIFIED AFOSR-TR-81-0324 NL

1 OF 2
AD A
097357



MAXIMUM ENTROPY IMAGE ESTIMATION

ANALYSIS OF IMAGE FORMATION WITH THINNED RANDOM ARRAYS

B.H. Soffer and Ryoichi Kikuchi

Hughes Research Laboratories
3011 Malibu Canyon Road
Malibu, CA 90265

February 1981

F49620-77-C-0052
Final Report
15 February 1978 through 31 August 1980

Sponsored by
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Building 410
Bolling Air Force Base, Washington, DC 20332

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFOSR-TR-81-0324	2. GOVT ACCESSION NO.: AD-A097357	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): MAXIMUM ENTROPY IMAGE ESTIMATION ANALYSIS OF IMAGE FORMATION WITH THINNED RANDOM ARRAYS.	5. TYPE OF REPORT & PERIOD COVERED: FINAL rept. 15 Feb '78-31 Aug '80 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Bernard H. Soffer and Ryoichi Kikuchi	8. CONTRACT OR GRANT NUMBER(s): F49620-77-C-0052
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Hughes Research Laboratories, 3011 Malibut Canyon Road, Malibu, CA 93065	10. PROGRAM ELEMENT PROJECT TASK AREA & WORK UNIT NUMBERS: 61102F 2305/B1
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research, Building 410, Bolling AFB, DC 20332	12. REPORT DATE: Feb 1981 13. NUMBER OF PAGES: 158
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS (of this report): UNCLASSIFIED 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum entropy, Image estimation, Signal estimation, Image processing, Signal Processing, Image restoration, Imaging, Antennas, Thinned arrays.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two related topics were studied in this three-year research program. The first was imaging behavior of thinned, or sparse arrays of small antennas, or apertures whatever the wavelength may be. The radiation properties of these arrays are known. We have made the first study of their imaging quality using computer simulation for coherent, partially coherent and incoherent radiation. Varying conditions of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

172600

signal to noise, phase aberration and monochromaticity were studied. For incoherent light, thinning down to 6% of the full aperture produced no noticeable degradation. The arrays are easy to design and degrade gracefully when large fractions of the elements fail. The result of this study can be found in the Interim Scientific Report.

The second topic studied, the one disucssed in this report, was the maximum entropy (ME) image reconstruction method. We formulate the problem from a fundamental point of view based on the physical statistics of the problem. The most probable object consistent with the measured signal image data and the noise is selected as the estimated object. The entropy is the logarithm of the probability. The constrained entropy is analogous, we note, to the free energy of statistical mechanics and we utilize this analogy to consider the problem of fluctuations, or noise in a natural way. The degree of confidence in the ME estimate is derived in a general manner by expanding the object probability distribution near the maximum as a multivariate Gaussian and making a principal axis transformation. The set of variances thus derived are projected back to the space of the object estimates as confidence regions. Several computer examples of the ME method in one and two dimensions are given.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SECTION 1

INTRODUCTION

This is the final report for a three-year research program on imaging with thinned sparse arrays of antennas or apertures. In the last year we have studied the foundations of the maximum entropy (ME) estimation method and its application to the restoration and super-resolution of images that have been degraded by undersampling and by the finite aperture of the sparse array. Extensive computer simulations were performed on one- and two-dimensional image signals to test the theoretical formulations of ME. This method of approach gives the results a certain generality in wavelength and physical scale that would have been extremely time consuming to obtain using the data from direct physical experimentation.

The simulations of imaging with sparse arrays provided important and surprising results. We summarize here the important points of our previous two-year study of these arrays. Further details can be found in the annual report for this contract (June 1979). Very high resolution imaging systems, regardless of wavelength, require large apertures with high resolution, low side lobe level, and high gain. These apertures are often considered to be technically and economically infeasible. We investigated the efficacy of using very sparse arrays of randomly placed small antenna elements in imaging systems. The properties of these arrays are established in antenna theory. What we have done is to make the first demonstration of their application in imaging systems. Computer simulation studies were made for coherent, partially coherent, and incoherent imaging. These were done using various conditions, including variable signal-to-noise ratio and phase aberration, but with monochromatic illumination. We showed that, for incoherent imaging, a large degree of thinning down to 6% of the number of full array elements (4096) produces image quality comparable to that produced by the full array. For broad-band

polychromatic incoherent imaging, even larger degrees of thinning, down to 3%, were similarly effective. The fracional degree of thinning allowable for incoherent imaging is in principle, inversely proportional to the square root of the number of elements in the full array, which is equivalent to the number of elements resolved in the full array. The arrays, which are easy to design, degrade gracefully when large fractions of the elements are removed. For coherent imaging applications, random arrays do not offer significant advantages over full arrays.

The body of this report describes the results of the last contract year's work on the ME estimation method. Both the theory and application are described along with the results of computer simulations on one- and two-dimensional images. The meaning and nature of the ME method is discussed in Section 2, where the meaning of entropy in this problem is clarified. A generalization is also presented that reconciles the two contending forms of entropy seen in the ME literature by showing them to be special limiting cases of a more general entropy, based on quantum statistics. The actual physical nature of the source and measurement apparatus determines the appropriate limiting form of entropy. The commonly used information theoretic entropy is shown to be one such limiting form.

An analogy is made between the ME problem and the free-energy minimization problem, familiar in statistical mechanics. This analogy allows fluctuation noise and signal-to-noise concepts to be introduced naturally into the theory and practice of the method via an effective noise temperature, if the noise or signal-to-noise is known a priori by independent measurement from an ensemble of measurements.

There is a closed-form solution to the ME problem in one dimension for the case of Gaussian statistics. In that situation, matrix inversion for image signals of small extent or recursion for larger ones can readily be used to solve the problem. No closed-form

solution exists for any other case, and iterative algorithms must be employed in digital computations. We studied two methods: the natural-iteration method, which guarantees convergence, and the much faster Newton-Ralphson method which does not. Examples of one- and two-dimensional and computer simulations of restoration and superresolution are given in Section 4. No approximation was needed for nonseparable two-dimensional point-spread functions.

A question rarely asked in ME estimation work is what degree of confidence should be ascribed to the estimates. We develop in Section 5 a general method for the calculation of the degree of confidence in the ME estimates, suitable for two dimensions and not restricted to Gaussian, independent statistics. This method is applied to a set of ME estimates consisting of estimated objects, all differing in the degree of superresolution and differing further in the degree of noise assumed. The multidimensional (49 dimensional in our example) nature of the results makes a two-dimensional graphical representation of the results difficult, but we have chosen to attempt it in order to supplement a purely algebraic statement.

As a corollary of our study and understanding of the entropy in the ME method, we noticed that many authors, while recognizing the importance of the Poisson statistical regime for imaging and communications, have incorrectly evaluated the probabilities and channel capacities for the Poisson distribution. They incorrectly used the wrong entropy or information expression for this case.

SECTION 2

MAXIMUM ENTROPY ESTIMATION: THE GENERAL POINT OF VIEW

Maximum Entropy has been used in various forms and for various ends in recent years. Applications ranging, for example, from radar filter formation to beam forming, from seismic spectral estimation, to astronomical imaging and even to economic forecasting, have all been enriched by this general technique. The particular techniques and philosophical points of view vary widely among the several disciplines, each working independently and publishing in specialized journals.

The interpretation we present, which we believe to be the most general and which includes others as special more limited cases, is based on two considerations. Any image, measured as signal, pattern, or spectrum, whatever the variables represent, is a necessarily degraded version of the true object. It is necessarily degraded because a real measurement system has limited bandwidth or finite point impulse response function, because the time sample is finite, and possibly because the measurements were undersampled. Furthermore, noise cannot be ignored.

Many different possible object patterns (whatever the object variables represent) all differ in detail, but can produce the same measured image pattern. This ambiguity can be resolved by applying the ME method. In our interpretation, the second consideration is that a probability is assigned to each possible pattern and the most probable pattern is chosen as the estimated or restored object. Patterns are assigned probabilities based on the physics and statistics of the immediate problem. The entropy is understood to mean the logarithm of the probability. So to find a maximum of the entropy is to find a maximum of the probability.

From this point of view ME is a method to estimate the true object (or its transform) by maximizing its entropy subject to the

measured image data constraints. In the next section we develop this idea in an analogy between ME and the well known, and much practiced statistical mechanical principle of the minimization of the free energy and derive some useful benefits from the analogy.

Here, some comments are in order about a much repeated "principle of maximum entropy" put forward by Jaynes which is often used to justify the ME method. In the particular form it is often quoted in ME, exploiting the relationship between information and entropy, it states that the ME estimate is the least prejudiced one bringing no additional information from tacit structures and assumptions. Any such implicit features would decrease the entropy from its maximum. We do not need to employ a new principle to formulate the ME problem. we maximize the probability and, following Boltzmann, we identify the logarithm of probability with the entropy.

An information theoretic point of view can enrich and sharpen our understanding of the relationship between the observer and the observed, the knower and the known, but it has sometimes been subject to misuse and misunderstanding. The information theoretic entropy of Shannon, -f log f, is often employed inappropriately. This form is appropriate only when there is no underlying a priori probability distribution departing from an equal a priori weighting. For example, it would be appropriate for the ubiquitous gaussian case but not so for the poisson case.

These considerations regarding the entropy expression have been developed at length in our paper "Maximum Entropy Image Restoration: I. The Entropy Expression," published in the Journal of the Optical Society of America and reprinted in Appendix A. The interested reader is referred there for a detailed exposition; only a summary is given here.

There are two schools employing two different forms of entropy in the ME problems they are solving. We call them the "log B" and "-B log B" schools, where B is the local or instantaneous brightness,

power, intensity, or their spectral counterparts. We derive these two expressions as limiting cases of a more general entropy forumula based on the underlying properties and statistics of the physical source and of the measurement process. In the case of photon or electromagnetic signaling or imaging, the Bose-Einstein statistics and in the case of electrons, the Fermi-Dirac statistics are employed. The n quantum mechanical particles comprising the brightness are distributed over z degrees of freedom as calcualted by these statistics. The number of degrees of freedom is a function of both the source and the measurement and estimation process. It can be understood in phase space in terms of the ratio of the sizes of the detection volume to the coherence volume of the particles. The number of degrees of freedom is an extension of the familiar idea of the time-bandwidth product to include the conjugate variables of space-reciprocal space as well. The entropy to be maximized is the logarithm of the probability as given by the physical statistics of the problem, following the original meaning of entropy as given by Boltzmann and Planck. The entropies log B and -B log B result in the limit $n \gg z \gg 1$ and $n \ll z$, respectively. When n is interpreted as an average $\bar{n}$ over an ensemble, we find in addition the log B expression when $n \gg z = 1$. The Burg form used in spectral estimation, for example, is log B. The distribution function for this case is the exponential in intensity, or Gaussian in complex amplitude. Shannon's entropy expression is shown to be, in the same way, a special case of the more general result and is appropriate for his special interest, namely, the $z = 1$ Gaussian complex amplitude case of equal a priori probabilities.

We show in Figure 1, by way of illustrating the $\bar{n}/z$ concept with a familiar example, a plot of the wavelength-temperature dependence of a blackbody for given average number of photons per mode (or per degree of freedom) $\bar{n}/z$.

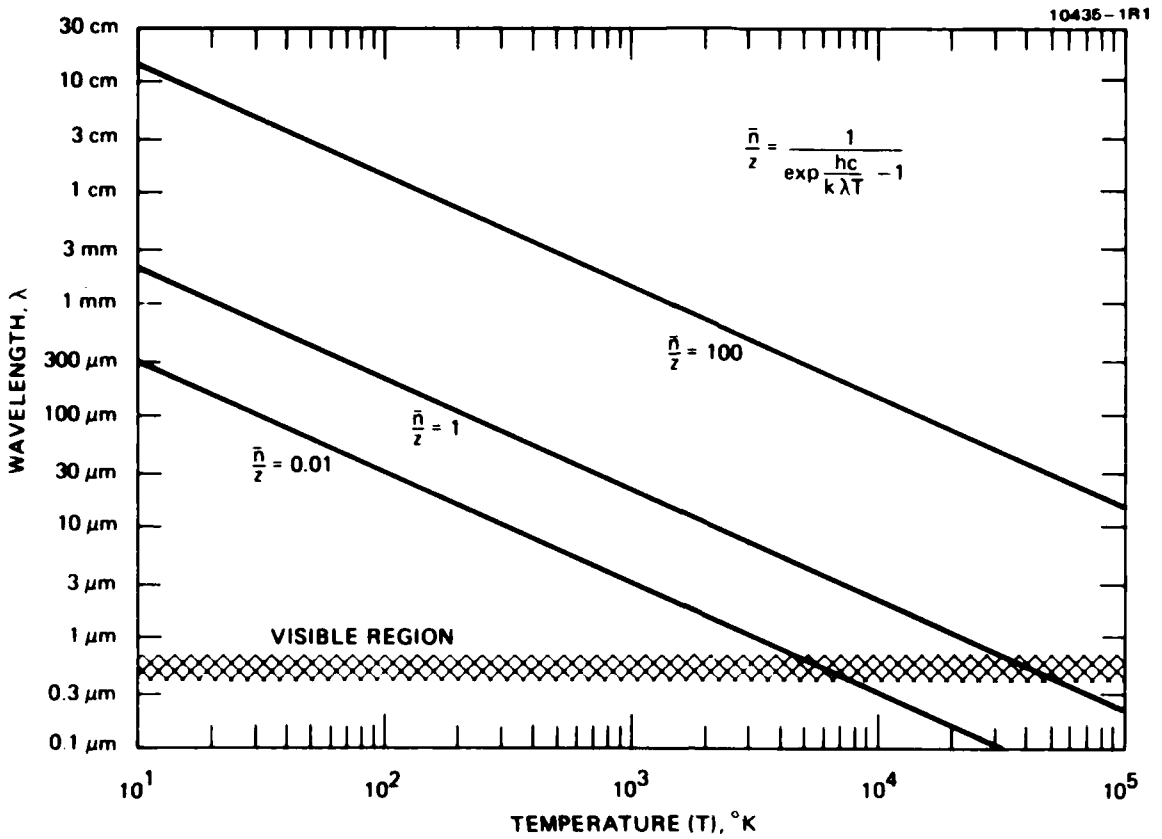


Figure 1. Wavelength-temperature dependence of a blackbody for given average numbers of photons per mode (or per degree of freedom), $\bar{n}/z$.

The ME method, in summary, may be viewed as a nonlinear inversion or deconvolution method for estimation from partial and noisy data. It can be used to extrapolate or interpolate in the spaces of the measured variables or their conjugates, as in Fourier transform space. Since it can achieve an inversion beyond the algebraically allowable spatial, temporal, or spectral limit of the measured data, it can be said to produce a superresolution. In the next sections we describe our method of solution, describe a useful thermodynamic analogy, and give some examples in one and two dimensions. Finally, we address the question of the degree of confidence in the estimates.

SECTION 3

THE ME METHOD AND ITS INTERPRETATION

In this section we describe our particular formulation of the ME method and present a useful thermodynamic interpretation in terms of the free energy. This interpretation permits us to deal with the problem of fluctuations or noise in a natural way. The Burg method and the relationship of ME to the Bayesian point of view is discussed.

A. FORMULATION OF THE ME PROBLEM

We begin with a set of measurement data in one or more dimensions, a known noise or signal-to-noise ratio (S/N), a known instrumental and transmission path point-spread-function for measured spatial variables, and impulse response function for measured temporal variables. If we are measuring or estimating spectra in the reciprocal or conjugate space of spatial or temporal frequency, we must then know the detailed behavior of the band-limited transfer function or time-limited sampling function of the instruments. The goal of the method will be to estimate a finer resolution version of the original measurement datum (or its spectra), or equivalently, an extrapolated version of the original measurement datum (or its spectra.) An example from image estimation will help to illustrate these ideas.

Suppose an unknown object {Ob} convoluted $\circledast$ with a known instrumental point spread function (PSF) is observed as an image signal {I} .

We may write

$$\{Ob\} \circledast \text{PSF} = \{I\} \quad . \tag{1}$$

Naturally, there are fewer independent variables or resolution elements (space-spatial frequency product elements) in the image than there are potentially in the object. There are many possible objects, that could produce the same image. There is a many-to-one mapping of the variables that might describe the object to the variables that describe the image signal. One ME problem is to invert Eq. 1 to get an estimate of the object $\{\hat{Ob}\}$ with some desired degree of resolution or superresolution. The carrot $\hat{}$ symbol denotes estimates.

Inversion is possible using linear methods, but only up to the band limit, resolution limit, or, the uncertainty limit. The problem is that of inverting singular matrices or dividing by zeros, or near zeros. All linear methods are basically equivalent in sharing this difficulty.

The classical analysis by Slepian and Pollock showed that at and beyond the resolution or uncertainty limit the eigenvalues fall rapidly to neglible near-zero values. Thus, any finite noise, whether physical noise in the signal or numerical noise in the digital computer, renders superresolution impossible by linear inversion schemes. As a nonlinear estimate method, ME diminishes these difficulties. Like other nonlinear methods, such as the iterative method of constrained positivity, ME can achieve superresolution estimates in the presence of noise. (We incorporate a positivity condition in our ME formalism in a natural non ad hoc way.)

Only the one-dimensional ME problem for the case of Gaussian statistics has a closed-form solution (discovered by Burg). This problem can be solved by direct matrix inversion, although recursion schemes are usually employed for speed and economy of calculation. In higher dimensional problems, such as two-dimensional pictorial imaging or synthetic aperture radar, no closed-form solution exists. Furthermore, for any statistics other than Gaussian, no closed-form solution exists. For example, for Poisson statistics, the problem must be solved numerically.

Our formulation of the ME problem is to assign a probability to each possible object, based on the physical statistics of the problem at hand, and choose the most probable object as the estimate. In practice, we do not enumerate the probability for each object that agrees with the constraints, but determine the most probable object by the methods of analysis. But, in principle, it could be done that way.

According to Boltzmann's identification of entropy S and probability P (see Appendix A) P being the normalized number of ways something can occur, the degeneracy

$$S = \log P \ . \tag{2}$$

We see that seeking the maximum probability object estimate is equivalent to seeking the maximum constrained entropy:

$$P = e^{+S \text{ constrained}} \ . \tag{3}$$

We write the constraint in the form of the squared difference between the given measured data and the estimate to be found. By subtraction, the constraint compares the calculated estimate with the original measured data. Before comparison, the two quantities must be in, or be transformed into, the same space and compared on the same support elements. For example, we may compare the superresolved object estimated on a fine mesh with the coarser mesh measured image by convoluting the estimate with the known point-spread function of the measurement technique. This procedure reduces the number of independent coordinates in the estimate and places it on the same coarse mesh of the image for comparison. Thus, the estimate object is degraded by the same measurement techniques that produced the actual image. This derived "image" is compared with the actual image in the constraint term. Continuing our example, we write

$$E = \sum_i \left(\{I\} - \sum_k PSF_{i,k} \circledast \{Ob\}\right)^2 \quad . \tag{4}$$

In the same way, we can constrain the estimated object to agree with an image measured in Fourier transform space. This situation would arise, for example, in the detection of the partial coherence visibility function as is done in radio telescopy. Here we would compare the transform of the estimated object, degraded by the transfer or filter function (the modulation transfer function (MTF)) of the measurement technique, with the actual measured image transform. Again, the estimated object is degraded in the same fashion as the actual image was. We would then have the constraint

$$E = \sum_k \left(\{I_k\} - \sum_i \left[\text{Fourier Transform } (i,k) \quad \{Ob_i\}\right] MTF_k\right)^2 \tag{5}$$

where {I} and {Ob} are in the reciprocal spaces of spatial frequency (f_x) and space (x), respectively. Other permutations of the transform relation are also possible in the constraint term. This constraint term may be viewed as an error term, or a noise or error-tolerant term. Wernicke and D'Addaria,[1] utilized it in their formulation of the ME problem with an arbitrarily chosen multiplier that was not clearly interpreted. The constrained entropy to be maximized may be written as

$$\underset{\text{constrained}}{\hat{S}(\{Ob\})} = \hat{S}(\{Ob\}) - \beta \hat{E}(\{Ob,I\}) \quad , \tag{6}$$

where β is a Lagrange multiplier that is chosen so that the constant of Eq. (4) is satisfied. The multiplier β determines the weight, or importance, of the constraint. In general we can make β dependent on i and multiply β_i inside of Σ_i in Eq. (4); however, for simplicity we

use a constant $\beta_i = \beta$. The constraint term E may be viewes as an energy term since it is a quadratic form; this is in analogy to physical systems which often have quadratic potential and kinetic energies. We chose this quadratic form to emphasize this analogy for a useful interpretation of the multiplier, and also for technical and mathematical reasons. Because we want the energy to become small in our extremum problem, any even function, such as the absolute value or E^n, where n is an integer, would do. But the quadratic form makes an algebraic and algorithmic simplification.

Maximizing the constrained entropy is the same as minimizing its negative, and we minimize a function which we call βF:

$$\beta F = \beta E - S \quad . \tag{7}$$

Equation 3 for the probability of the object may now be written

$$P = e^{-\beta F} \quad . \tag{8}$$

Since maximizing the probability is the same as maximizing its monotonic mapping by the logarithm, we minimize βF (Eq. 7) directly:

$$\text{MIN } \beta F\{\hat{Ob}\} = \text{MIN}(\beta E - S) = \text{MIN } \left[\beta(\{I\} - PSF \circledast \{\hat{Ob}\})^2 - S\right]. \tag{9}$$

To illustrate more explicitly, we can choose an explicit form for the entropy such as $S = +\log\{\hat{Ob}\}$ (suitable for Gaussian amplitude statistics), or $S = -\{\hat{Ob}\}\log\{\hat{Ob}\}$ (suitable for Poisson statistics as we show in Appendix B). The convolutional PSF might be Gaussian or $(\frac{\sin x}{x})^2$, depending on the details of the measurement conditions. (We choose the $(\sin^2 x)/x^2$ form in our later examples.)

To minimize Eq. (2), we set the partial derivatives to zero,

$$\frac{\partial F}{\partial \{\hat{Ob}\}} = 0 \quad , \tag{10}$$

and solve for $\{\hat{Ob}\}$.

These equations can be solved iteratively by a variety of techniques. We have used a "natural iteration technique"[2] developed by Kikuchi, which guarantees convergence irrespective of the initial guess, and the Newton-Raphson technique[3].

B. A THERMODYNAMIC ANALOGY

When we examine Eq. 7, we note that it has the form of the function that is minimized in the canonical ensemble treatment of statistical mechanics. Using this analogy, we can go one step further and call F (defined in Eq. 7) the free energy. In our interpretation, we identify the constraint energy, Eq 4 or 5, with the internal energy of the system, and β with 1/T, the reciprocal temperature. Therefore, in our formulation, of the ME problem should really be called a Minimum-Free-Energy (MFE) problem. One may call the entire expression, Eq 7, entropy or free energy, depending on whether one views the system as closed or open, and whether the problem is viewed from the microcannonical or cannonical point of view. In any case, the same function is to be minimized. The important point is that in either case the multiplier β can be associated with the concept of temperature and therefore fluctuation or noise as we shall see below.

This analogy immediately indicates what value β should take to satisfy the constraints. From statistical mechanics we know that the system takes the state of the lowest energy at T = 0. This means that the value of β that satisfies the constraint is

infinity. When $\beta = 0$ (i.e., $T \to \infty$), the entropy predominates in Eq 7; thus the object viables correspond to those values that maximize S without the constraints, thereby producing a flat featureless distribution for the object. As β increases, the E term increases its contribution and is not at its lowest zero value.

Analogy with the statistical mechanics enables us to use the relation

$$\frac{\partial F}{\partial T} = - S \quad . \tag{11}$$

This relation can be used as a check of numerical computation. Equation 11 can be used as follows. If we can expand F near $T = 0$ ($\beta = \infty$) as

$$F = F_o - TF_1 - T^2F_2 \ . \ . \ , \tag{12}$$

then

$$S = - \frac{\partial F}{\partial T} = F_1 + 2TF_2 + \ . \ . \ . \tag{13}$$

and

$$E = F + TS = F_o + T^2F_2 + \ . \ . \ . \ . \tag{14}$$

Because we know $E = 0$ at $T = 0$, Eq. 14 leads to $F_o = 0$. Equation 13 shows that F is the residual entropy, i.e., the entropy value at $T = 0$. The behavior of F and S near $T = 0$ is shown in Figure 2.

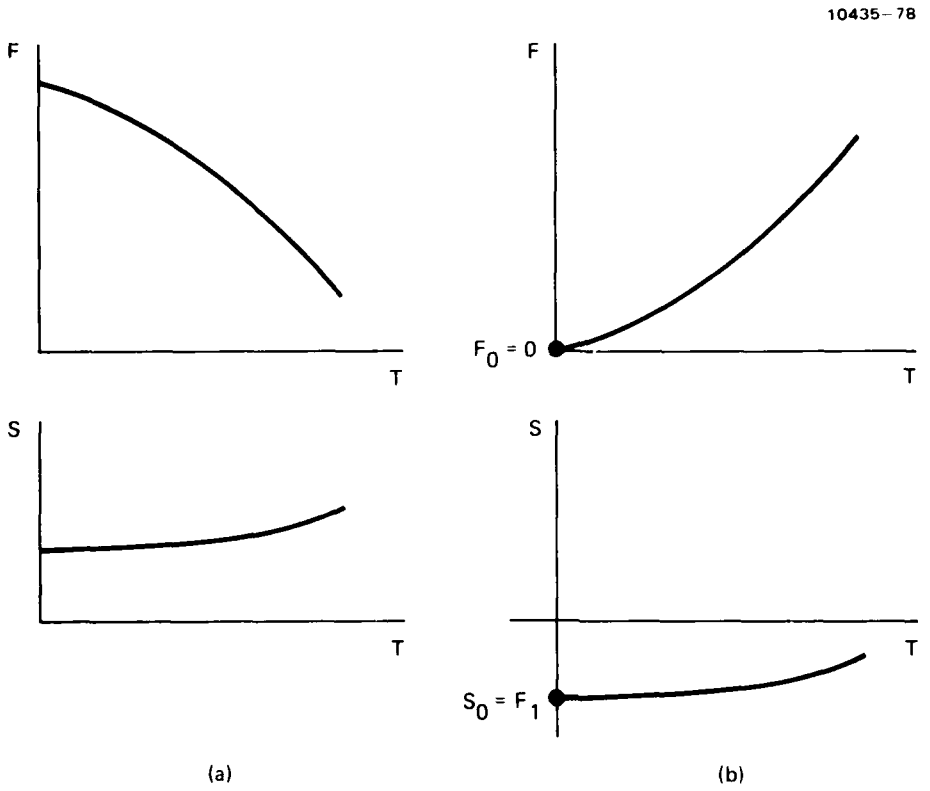


Figure 2. Behavior of free energy and entropy at low temperatures. (a) Normal behavior; (b) Apparent behavior in our formulation due to suppression of a large constant positive entropy term.

For the region of β we used in one of our examples, the expansion (Eq. 12) does not hold exactly, and the relation was

$$F = - T^{1+\alpha} F_1 \tag{15}$$

with a small α. In this region, for example, it was numerically verified that

$$S = (1+\alpha)T^{\alpha}F_1$$

(16)

and

$$E = \alpha T^{1+\alpha}F_1$$

hold with $\alpha = 0.05$. The relations in Eq 16 are derived from the thermodynamic relations, Eq. 11 and 14.

C. FLUCTUATIONS OR NOISE

Our thermodynamic analogy gives a further insight into the problem of the meaning of the energy constraint term and its multiplyer β. Wernecke and D'Addario,[1] noted that if the measured data values in the constraint are independent zero mean, random variables with known a priori variances σ_i^2, the constraint E could be written

$$E = \sum_i^M \left(\frac{\text{Estimated}_i - \text{Measured}_i}{\sigma_i} \right)^2 = M \quad , \tag{17}$$

the equality holding provided M is large. The computed trial solutions for various multiplyers β, to find the one that permits Eq. 17 to be satisfied with "sufficient" accuracy. They note that it might not be possible to satisfy Eq. 17 no matter how large β if the measurements are too noisy or the variances $\{\sigma_i^2\}$ are assumed to be unrealistically small. They give an estimate for a trial guess for β from the case of a single measurement of the total object intensity m_1 by setting $\partial F/\partial Ob = 0$ (Eq. 10), and setting the difference between measured and estimated values equal to σ_1,

$$\beta = \frac{1}{2\left(1 + \frac{m_1}{\sigma_1}\right)} , \qquad (18)$$

which varies from zero to a maximum of one-half. They increase β to satisfy "increasing accuracy" in Eq. 17, but no further rule or interpretation is given.

Gull and Daniell,[4] seemingly unaware or Wernecke and D'Addario's prior work[1], similarly assume the data to have Gaussian errors and note that then the term E (Eq. 17) is a χ^2 distributed statistic and that the expected value of χ^2 is equal to the number of data values M. They automatically increase β in their iterative solution until this equality is achieved. No interpretation of β is offered there to relate it explicitly to the noise.

In thermodynamics the fluctuation of the energy in a canonical ensemble is given by

$$\overline{\delta E^2} = k\, T^2 C_V = k \frac{C_V}{\beta^2} , \qquad (19)$$

where we will, in our analogy, put Boltzmann's constant k = 1. The specific heat C_V is given by

$$C_V = \frac{\partial \bar{E}}{\partial T} = T \frac{\partial \bar{S}}{\partial T} . \qquad (20)$$

These two relations are general and do not depend on the assumption of Gaussian independent error or noise fluctuation statistics, or the validity of the equipartitioning of energy for the problem at hand.

If equipartition were to hold for the ME problem, we would have a simple theory for the specific heat and the fluctuations. Equipartition would indicate that the fluctuation or noise power is distributed uniformly among all the pixels, or mesh points, with $1/(2\beta)$ in each pixel, each pixel being a degree of freedom in our present equipartition analogy. Equipartition would hold if the quadratic energy term were the only factor in the probability or free energy expression (Eq. 8):

$$P = e^{-\beta \sum_{i}^{M} \left(\Delta_i\right)^2} , \tag{21}$$

where Δ is the difference between estimate and observation. We could set

$$\beta \sum \left(\Delta_i\right)^2 = 1 \tag{22}$$

and have a simple relationship between β and the noise by relating Δ_i to the noise. We would employ a single β for simplicity rather than separate multipliers β_i. But the nontrivial entropy S in Eq. 8 spoils the relation (Eq. 21) for equipartioning except for special asymptotic limiting cases:

$$P = e^{-\beta \sum_{i}^{M} \left(\Delta_i\right)^2} \cdot e^{+S} . \tag{23}$$

In general, we must numerically calculate C_V, which is β dependent, and then use Eq. 19 to relate the value of β to the fluctuation or noise $\overline{\delta E}$. The numerical calculations are required because the series expansions do not hold exactly as we pointed out in Eq. 15 and we have not yet developed a theory of the specific heat for this problem. We emphasize that the relation, Eq. 19, does not depend on the noise being of the nature of Gaussian-independent fluctuations, but is very general.

Some comments may clarify the meaning of the energy fluctuation $\overline{\delta E^2}$ (Eq. 19). It is important to note that although the individual terms within the energy constraint expression, Eqs. 4 and 5, ordinarily have the meaning of physical energy (power, brightness, intensity, etc. or their spectral counterparts), we do not refer to these physical energies in the phrase "energy-fluctuation." Rather, it is our artificial analogy with the ubiquitous quadratic energy form such as occurs in the simple harmonic oscillator that permits us to call the entire constraint term an "energy" in our analogy. It is the fluctuation of of the entire term that is determined by Eq. 19.

If we write the measured quantities m_k and the estimated quantities on the same grid as n_k we may write Eq. 19 as follows. First we write δE noting that averages are done on n_k but not on the given m_k:

$$\delta E = E - \bar{E} = \sum_k \left[\left(n_k - m_k \right)^2 - \overline{\left(n_k - m_k \right)^2} \right] \tag{24}$$

$$= \sum_k \left[n_k - \overline{n_k}^2 - 2 \left(n_k - \bar{n}_k \right) m_k \right] .$$

We then must square and average to get

$$\delta E^2 = \overline{\left\{ \sum_k \left[n_k - n_k^2 - 2 \left(n_k - \bar{n}_k \right) m_k \right] \right\}^2} . \tag{25}$$

This complicated collection of higher moments can be simplified when the statistical distribution of the estimated object (n) is known. For example, if it is known to be negative-binomially distributed (see Appendix B) the second moment may be calculated as

$$\overline{\left(\Delta n \right)^2} = \bar{n} \left(1 + \frac{\bar{n}}{z} \right) . \tag{26}$$

Higher moments for this and other distributions may thus be calculated separately to simplify Eq. 24. Arbitrary definitions for the fluctuation such as

$$\frac{\overline{\delta E^2}}{\overline{E^2}}$$

or the root-mean-square fluctuation

$$\sqrt{\frac{\overline{\delta E^2}}{\overline{E^2}}} \tag{27}$$

may be usefully employed as well to simplify the results.

D. THE BURG METHOD

The Burg, or one-dimensional Gaussian statistics ME method and related algorithms has been shown to be equivalent to an all-pole network model. These ideas have inflitrated estimation theory from the discipline of control theory. Statistics other than Gaussian do not necessarily produce all-pole solutions. There are those who hold that if a process does not have such a network model representation, there is no physical reality involved. (See Appendix B.)

In the one-dimensional Gaussian-Burg problem, as generally practiced, a fixed number of autocorrelations or, alternatively, lags in the temporal data string is chosen; this number is also the number of poles that are determined in the spectrum. The location and magnitude of these poles determine the position and strengths of the spectral peaks. Where they fall, how close they are to each other, and how strong they are, are all the result of the calculation. In our formulation of the Gaussian problem we choose the fineness of the mesh upon which the estimates are constructed and upon which they are constrained to stand. Only the strength of the estimates at these mesh points is calculated.

E. BAYESIAN INTERPRETATION OF THE ME METHOD

The Bayesian approach may be viewed as a method for computing the conditional probability or probability distribution of a state of nature (or a cause) given the measured data (or an effect). If we write $P_A(B)$ for the conditional probability or probability distribution of B given A, then Bayes' theorem may be written

$$P_A(B) = \frac{P(B)\ P_B(A)}{P(A)} \quad . \tag{28}$$

In the ME problem we seek to determine the most probable object given an image. That is, the maximum of $P_I(Ob)$,

$$P_I(Ob) = \frac{P(Ob)\ P_{ob}(I)}{P(I)} \quad . \tag{29}$$

Now the probability distribution of images given an object $P_{ob}(I)$ is determined by the instrumental and transmission channel characteristics, including noise, by the PSF for example. The probability distribution of the images P(I) is a constant with respect to the maximization over objects. Wernecke and D'Addario[1] questioned the possible relationship between the Bayesian or maximum a posteori probability method (MAP) and ME, but commented that P(ob) was a stumbling block and that to use MAP one must make a model for the statistics of the object P(ob). From our point of view, the distribution P(Ob) is given by the physical statistics of the problem. For example for photon signals, the Bose-Einstein statistics apply (as discussed in Appendix B). With these identifications and interpretations, the maximization of $P_I(Ob)$ for the most probable object done in the ME method may be viewed as a maximization of Bayes' equation, a MAP method.

SECTION 4

EXAMPLES OF ME ESTIMATION

We illustrate the ME method with some examples of one- and two-dimensional problems and modeling both Gaussian and Poisson statistics. Although we conceived of these problems as photon imaging problems to address the issue of imaging with thin, sparse arrays, it will be obvious that by changing the names and meanings of the variables a wider variety of problems is implicitly included.

A. ONE-DIMENSIONAL EXAMPLE

In the first example, a one-dimensional photon signal is received by an array of antennas or apertures. The signal is supposedly known to be the far-field radiation pattern of the unknown object and is the mutual or partial coherence function; i.e., by the Zernike-Van Cittert theorem, it is closely the Fourier transform of the unknown intensity distribution pattern of the object. The array samples a small discrete subset of object transform values and the ME method estimates the object. The constraint energy term E in this example takes the squared difference between the measured values and the estimated K transform values, calculated from the estimated normalized object intensities $\{\hat{p}_i\}$ on a fine mesh labeled i ($1 \leq i \leq i\ \max$):

$$E\{\hat{p}_i\} = \sum_{k=1}^{k} \left(\sum_{\{i\}} \hat{p}_i \cos\left(2\pi f^{(k)} \cdot x_i\right) - H_k^{\text{measured}} \right)^2 \qquad (30)$$

The spatial frequencies $f^{(k)}$ in the incoherent case are a measure of the separation of pairs of elements in the array. The number of pairs K will be less than the number of mesh points i max, so we will be asking for a superresolution. We use a cosine transform here and create an artificial extension of the object to make an even function. This was done for convenience.

Notice that we are constraining in transform space but estimating $\{\hat{p}_i\}$ in object space in this example. An additional constraint is placed on the overall object intensity so we will estimate only the relative shape of the object,

$$\sum_i \hat{p}_i = 1 \quad . \tag{31}$$

The free energy times β becomes

$$\beta F = \beta E\{\hat{p}_i\} + \sum_i \left(\hat{p}_i \ell n\, \hat{p}_i - \hat{p}_i\right) + 1 + \lambda \left(1 - \sum_i \hat{p}_i\right) \quad . \tag{32}$$

This function is to be minimized with the Lagrange multipliers β and λ for the two constraints. The minimum is obtained when

$$\frac{\partial \beta F}{\partial \hat{p}_i} = \ell n\, \hat{p}_i - \lambda + \beta \frac{\partial E\{p_i\}}{\partial \hat{p}_i} = 0 \quad . \tag{33}$$

is satisfied for all $\hat{p}$. This simultaneous set of equations to be solved for p_i is treated by a special kind of iteration scheme.[2]

We can guarantee the desired positivity of the $\{\hat{p}_i\}$ by writing it as $\hat{p} = \rho^{q_i}$ in our formulation. This is a natural way that does not simply reject or set to zero negative values as sometimes reported in other nonlinear schemes.

In this example we estimate $\hat{p}_i$ on a support of imax = 50 points. We observe the first seven of fifty spatial frequencies. In our simulation this would represent a potential seven-fold superresolution at large values of β if the data allow it. To further illustrate the possibilities of the method, the sixth of these seven spatial frequency components was suppressed, simulating a particular sparse array. The diffraction, band limited, or resolution limited object estimate (further limited by the missing component) that was obtained by Fourier transformation of the six components of given measured image data, is shown in Figure 3.

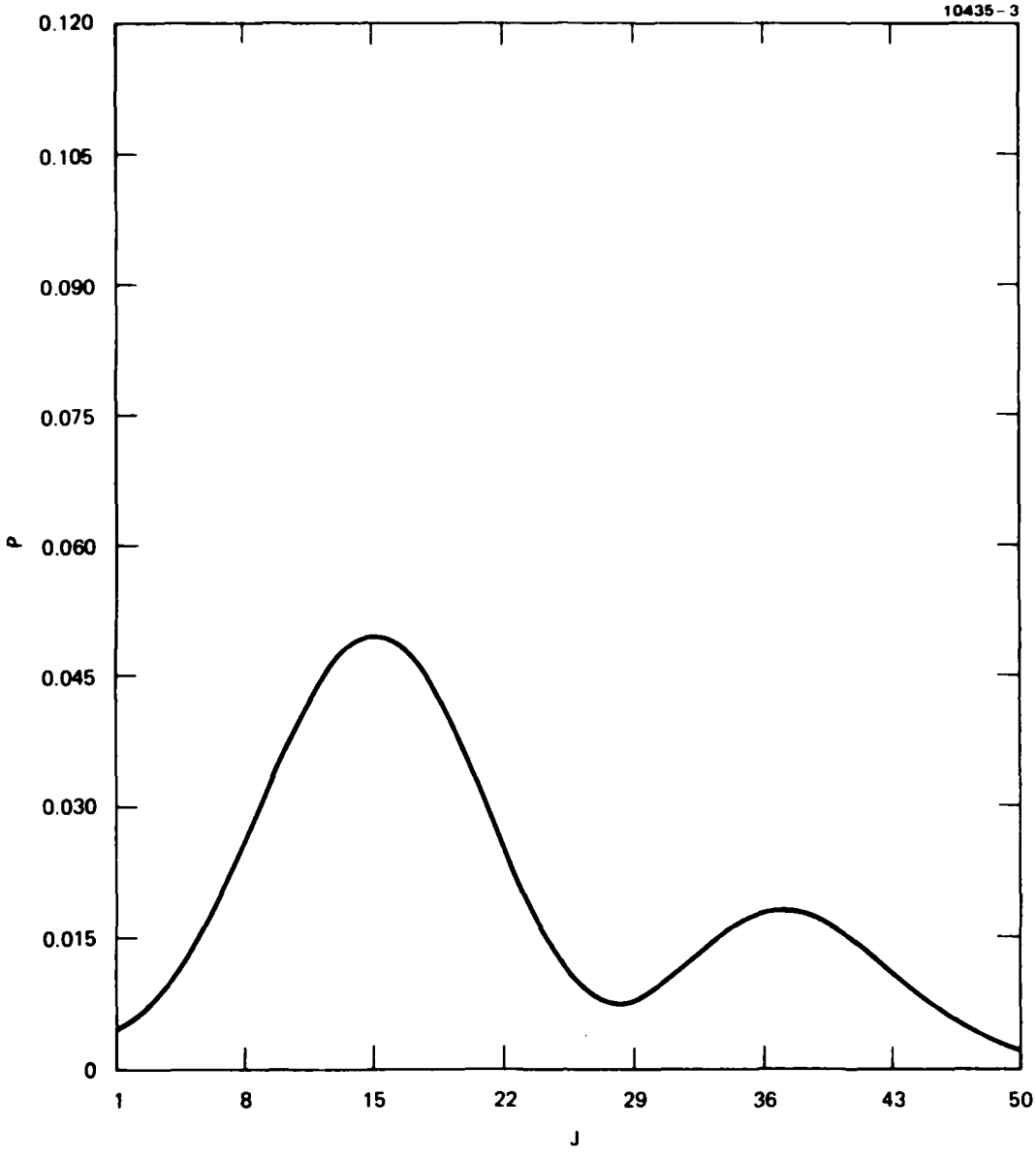


Figure 3. The diffraction-limited object estimate (Fourier transform of the observed or measured image data) further degraded by one missing Fourier component.

A series of fifty dimensional estimates for the objects $\{\hat{p}i\}$ is shown in Figure 4 for seven different values of β ranging from one-half to 120. It can be seen that smaller values of β, representing a lower known confidence in the accuracy of the data or equivalent by a lower S/N, produce a smoother, flatter estimate, while large values of β produce sharper, more detailed estimates. The diffraction object estimate falls somewhere between the curves for $\beta = 1/2$ and $\beta = 1$, but there is no curve of definite β that can reproduce the diffraction pattern. This behavior can be seen more clearly in Figure 5. A curious behavior can be noted near the peak below J = 15 in Figure 4. The $\beta = 120$ estimate takes a slightly lower peak value and the entire peak structure moves slightly to the left. Why this is a free energetically favorable situation is not obvious. This example is not a severe test of the ME method as there is not much structural detail at the fineness of the support of 50 points. The next example will show diffraction-limited behavior more explicitly.

B. TWO-DIMENSIONAL EXAMPLES

In the two-dimensional example we now turn to, the energy constraint is written entirely in the spaces of the object and the image, unlike the previous example where it was written in the transform spaces, of object and image. In these examples we use the Log B formulation of ME appropriate for Gaussian amplitude or exponential intensity statistics. We may imagine, in these examples, that the measured data are given directly as an image in real space or that it has been transformed to image space from measurements in Fourier space. The form of the constraint was described in Eq. 2. We chose a convolutional point spread function of the measurement to be of the form

$$\frac{\sin^2\left(x-x'\right)}{\left(x-x'\right)^2}\ \frac{\sin^2\left(y-y'\right)}{\left(y-y'\right)^2}$$

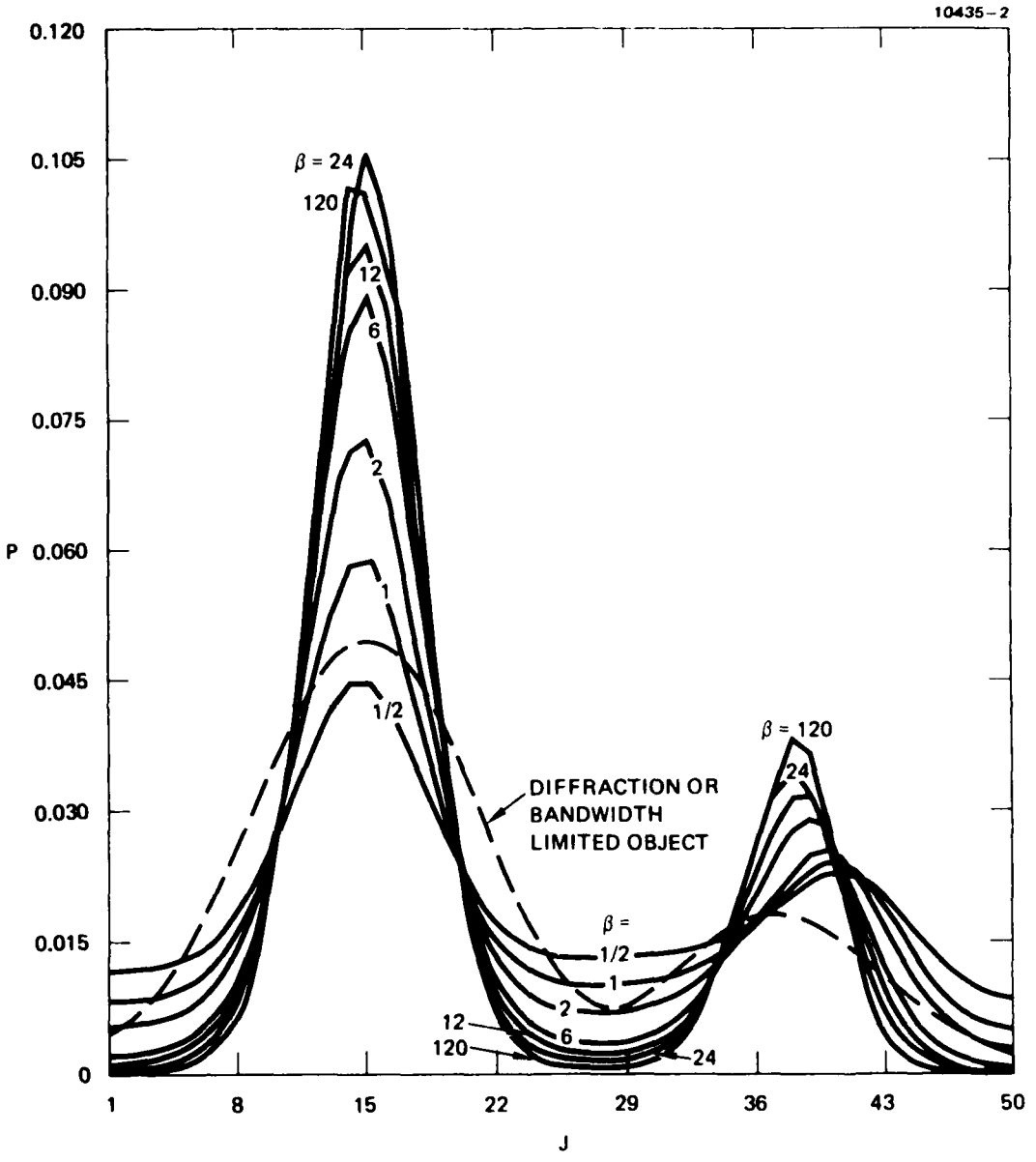


Figure 4. A family of ME object estimates for several values of the parameter β, compared with the diffraction object estimate.

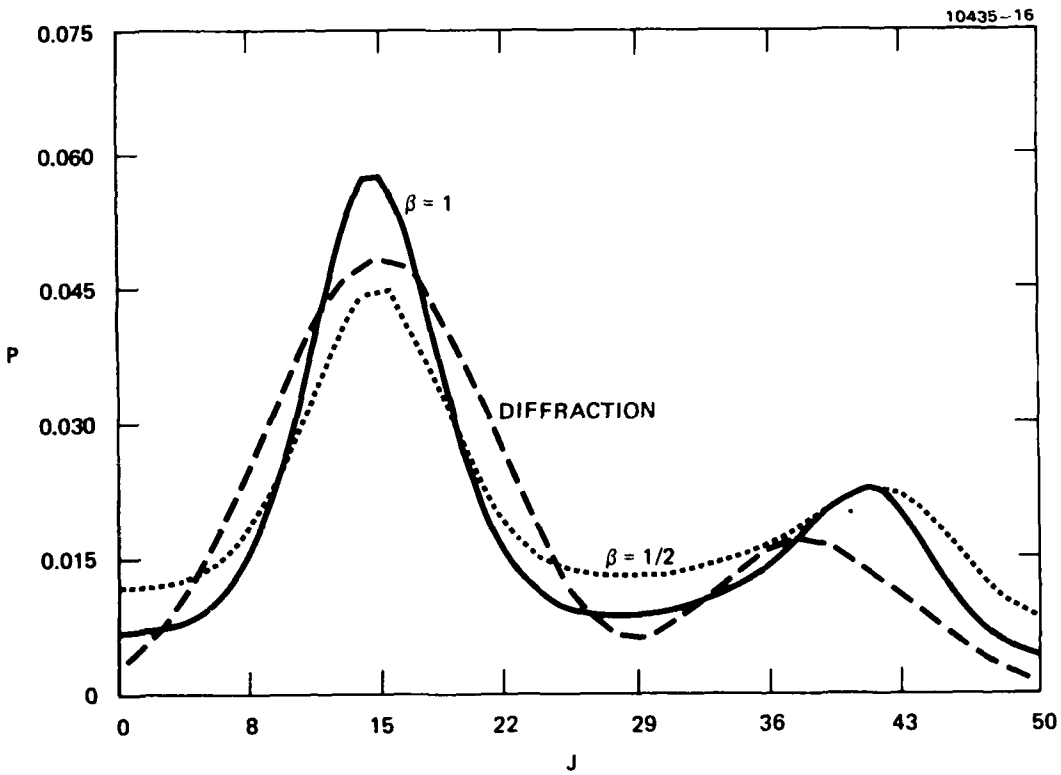


Figure 5. There is no ME object estimate for any β that can reproduce the diffraction object estimate.

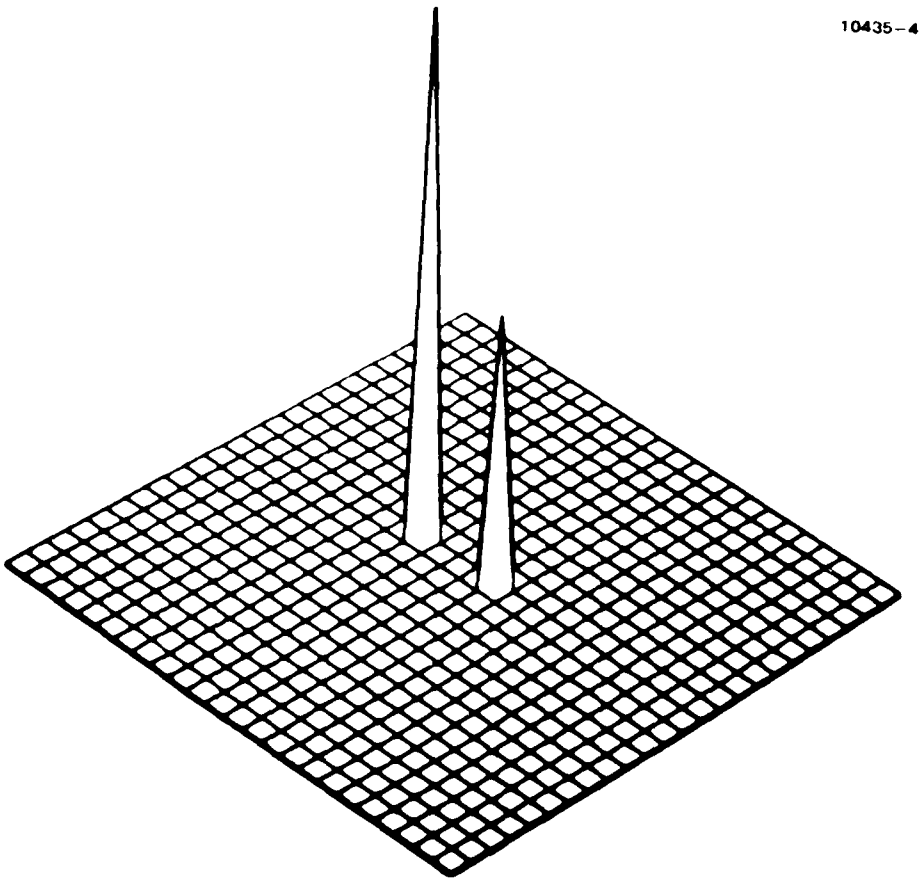


Figure 6. Two-dimensional test object at the resolution separation of the aperture.

representing diffraction spreading by a limiting square aperture. The scale of diffraction spreading was chosen to be a factor of four in each dimension. That is, the number of independent variables needed to completely describe the image was 4 x 4 = 16 times less than required to describe the object from which it was mapped. The image can be completely and uniquely represented on a grid 4 x 4 times coarser than the object.

In Figure 6 we show a two-dimensional test object of two isolated spikes supported on single-grid locations separated by three empty grid cells. Note that they are on a repeat period of four grid cells. The heights of the spikes are in the ratio of 255 to 128, and the base of the figure is unit height. The 4 x 4 times degraded diffraction-limited image is shown in Figure 7 on an appropriately reduced coarse support grid. If the two spikes of the object in Figure 6 were of the same height, then the image in Figure 7 would be said to have "resolved" the spikes according to Rayleigh's criterion, but as the spikes were of different heights, they are just unresolved. For convenience, both of visualization and of computation, the image is redrawn on the same fine grid of the object in Figure 8. These interpolated image values are redundant and contain no additional information. They are uniquely determined by the values on the coarse grid in Figure 7.

The ME estimate of the object was made on a 20 x 20 section of this image on the fine grid so the degree of superresolution of the image was a factor of 4 x 4. The result for the case $\beta = 10^4$ is shown in Figure 9. The two peaks are clearly resolved and some widening at their bases can be seen.

The iteration process was terminated by testing for when the sum of the absolute differences between the estimated values in two successive iterations was less than or equal to 10^{-5}, a number chosen arbitrarily. Some arbitrary choice is needed to terminate the process of ever diminishing returns. At this test value the peak heights were 254 and 96. In the natural iteration method[2] used in these examples, the free energy necessarily decreases monotonically with iteration for

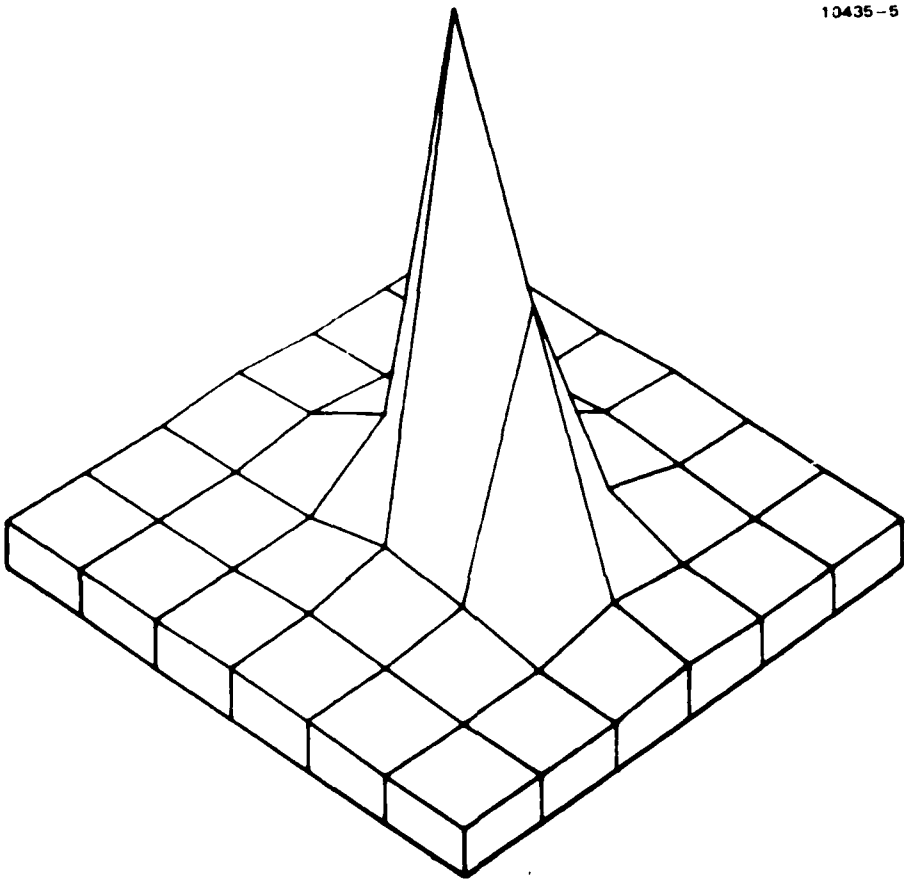


Figure 7. Just unresolved image of the test object of Figure 6.

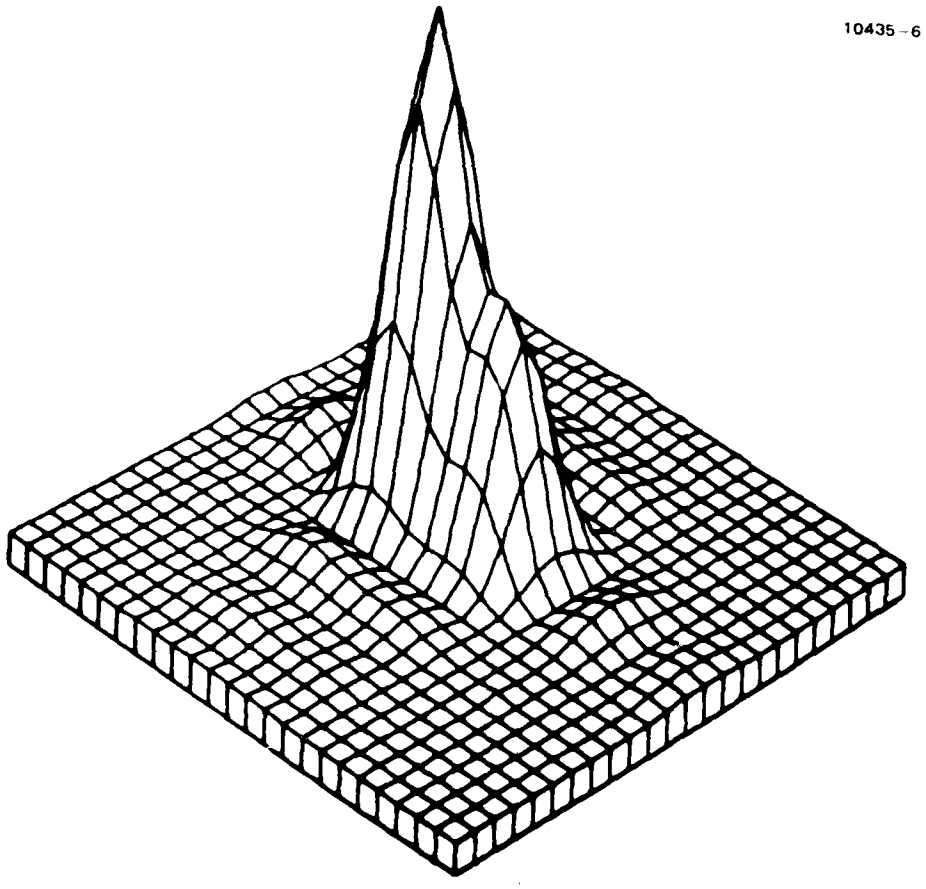


Figure 8. Fine mesh representation of the image in Figure 7.

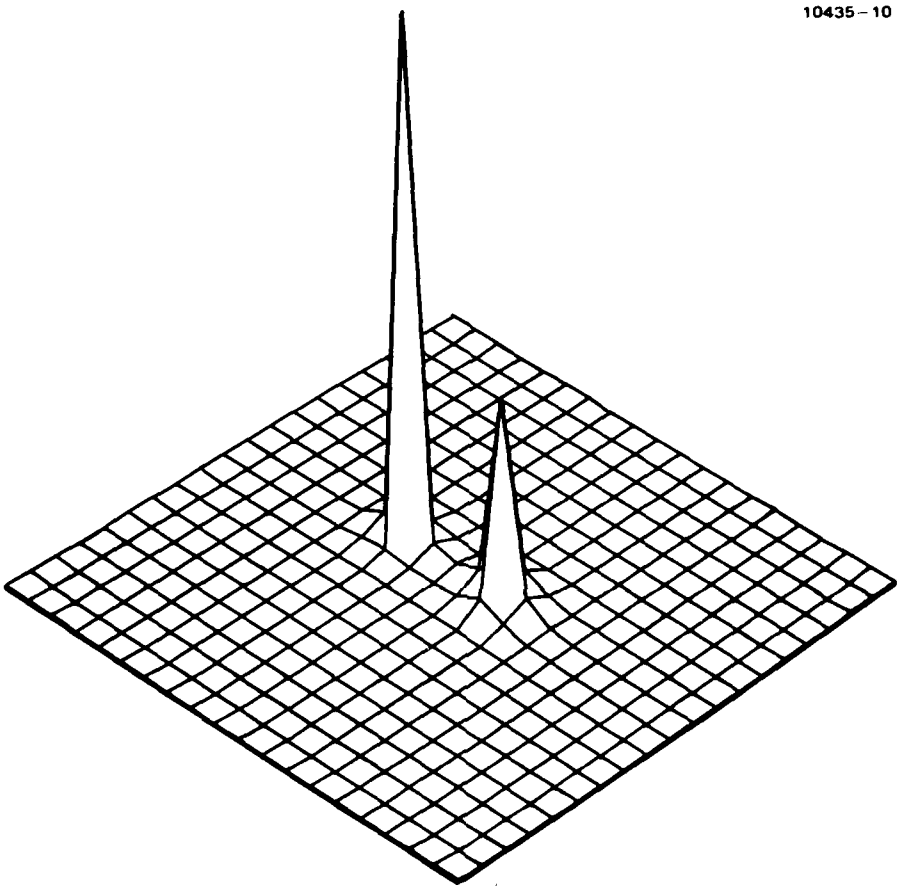


Figure 9. ME object estimate of the just unresolved image of Figure 7 (or of Figure 8)

all β values. The test described above was found to decrease exponentially but only for β below a certain value, depending on the details of the problem. Above that β value, the convergence was impractically slow and erratic. The Newton-Raphson method does not guarantee convergence from an arbitrary initial guess and also fails to converge for values of β greater than a certain maximum value.

The second two-dimensional example we present is similar to the previous one, except that here the image is derived from the test object, shown in Figure 10, whose two spikes are at half the previous separation, i.e., at half the resolution separation of the aperture. The unresolved image is shown in Figure 11 on the appropriate coarse grid. Figure 12 shows the fine mesh representation of that image. The ME estimate of the object is shown in Figure 13 for $\beta = 10^5$. The unresolved peaks in the image are now clearly resolved and the peak values are 255 and 126. A saddle point of height 40 sits between the peaks, and there is a slight width at their base. The displacement of the peaks in this figure is merely a result of an improper plotting program instruction and should be ignored.

The last two-dimensional example we present is constructed from binary black and white alphabetical object shown in Figure 14(a), constructed on a 20 x 20 grid. The four by four-fold diffraction image is seen in Figure 14(b). The 400-dimensional ME estimate of the object is shown in Figure 14(c) for the case of $\beta = 10^7$. The convergence was so slow and erratic at this high value of β that Figure 14(c) shows rather unconverged results; that is, the fluctuations between successive iterations were still large. However, the superresolution was achieved. But the gray level or intensity values in the ME estimate are still uncertain in Figure 14(c).

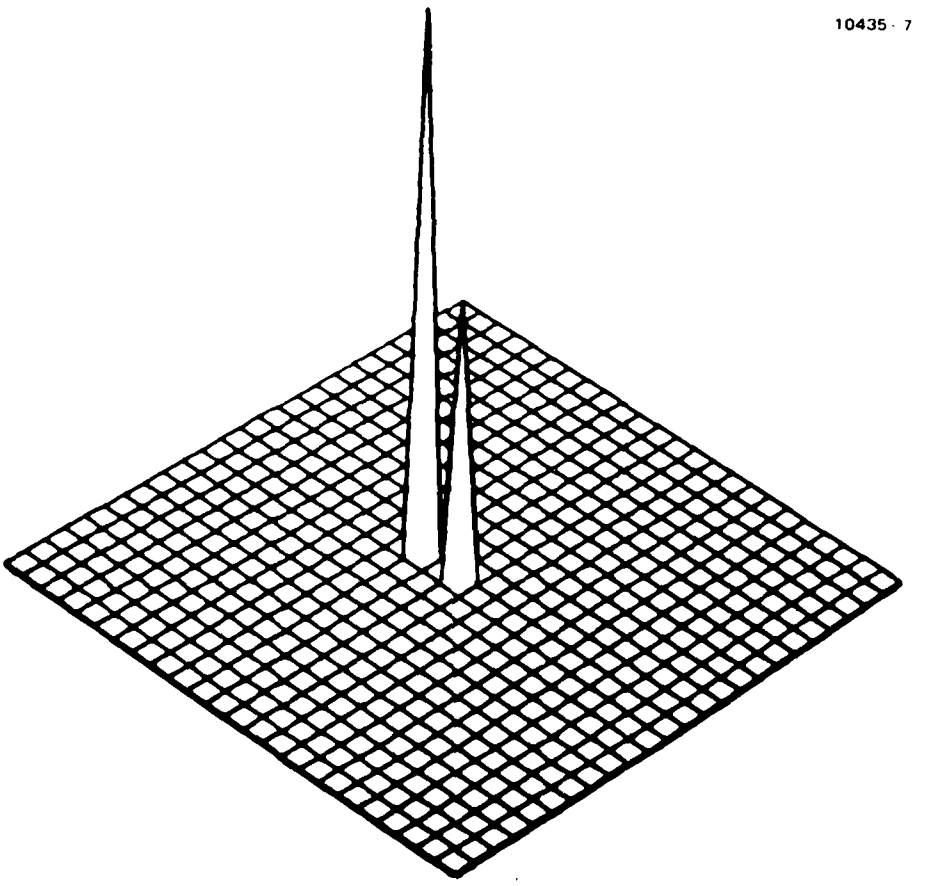


Figure 10. Test object at half the resolution separation of the aperture.

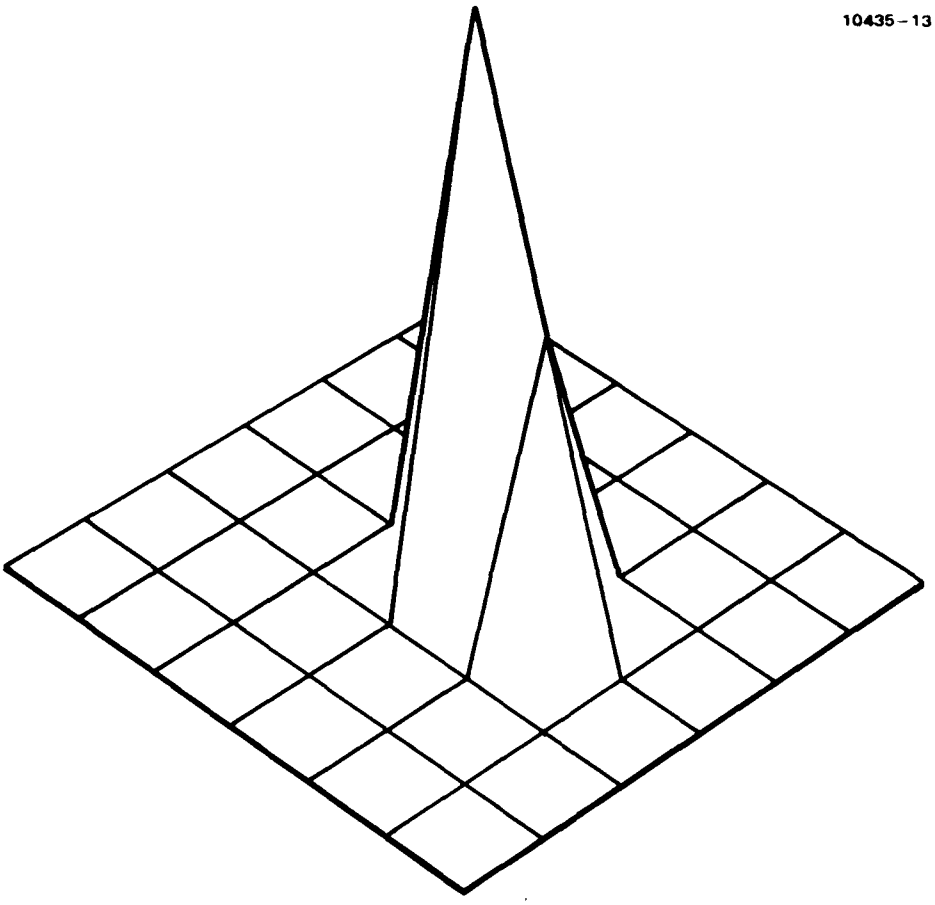


Figure 11. Unresolved image of the test object of Figure 10.

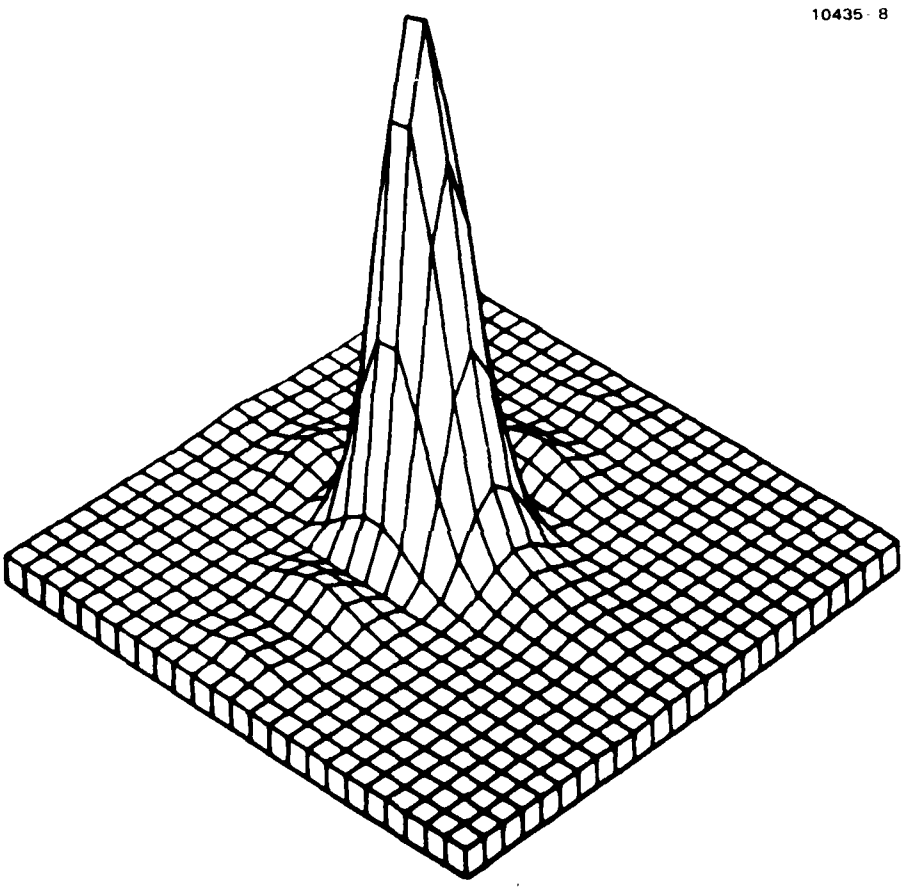


Figure 12. Fine mesh representation of the image in figure 11.

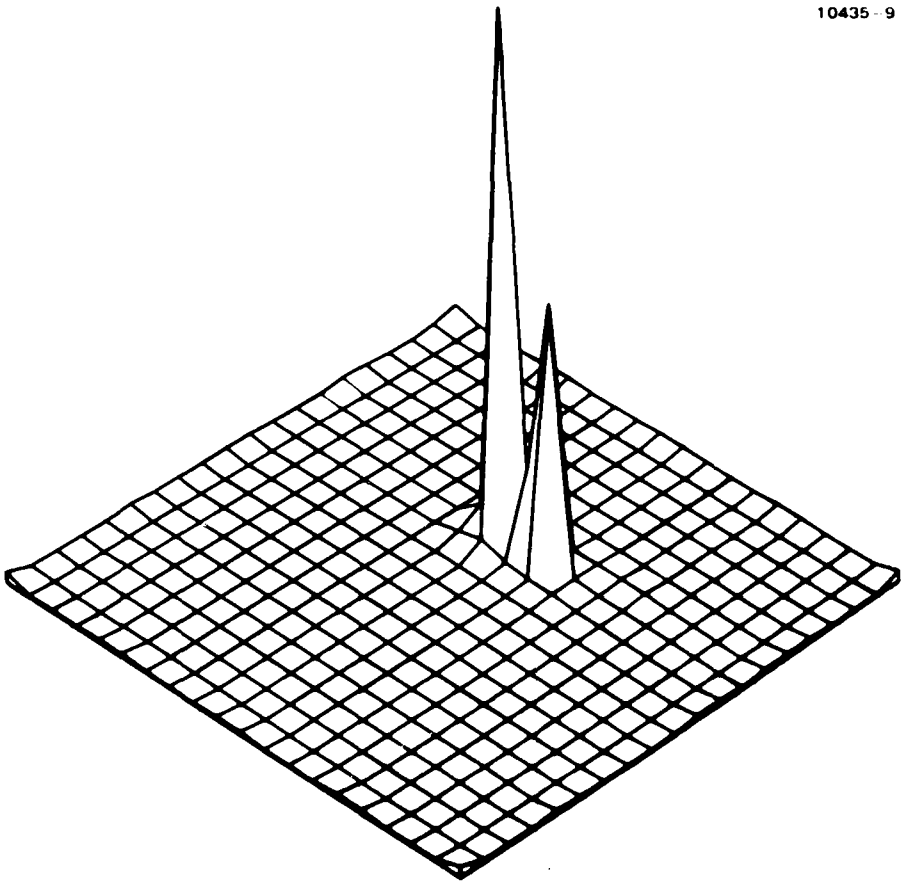
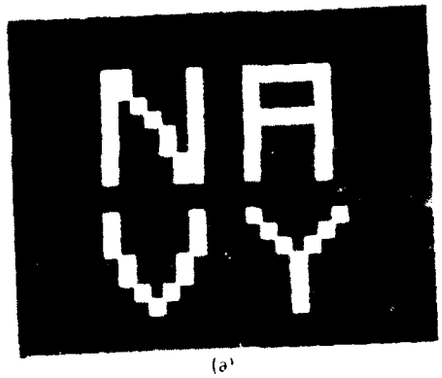
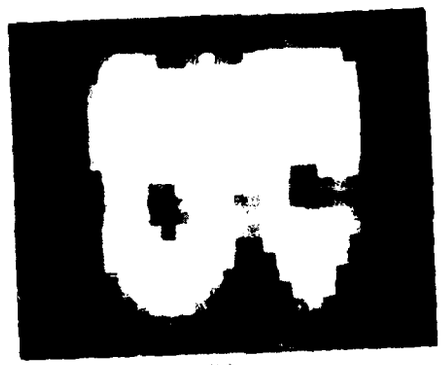


Figure 13. ME object estimate of the unresolved image of Figure 11 (or Figure 12).



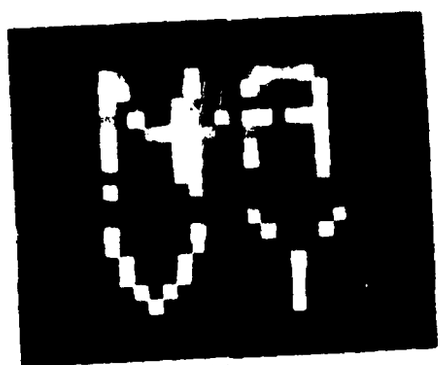
(a)

Figure 14(a). Sharp aperture test object.



(b)

Figure 14(b). Diffraction limited image.



(c)

Figure 14(c). NA object criteria.

SECTION 5

THE ME CONFIDENCE OR RELIABILITY ESTIMATE

When an object is estimated using the ME method, the question naturally arises concerning how much confidence is to be placed in the estimate. In this section we describe the theoretical formulation for the calculation of the confidence, and we give some examples for a variety of measurement conditions.

A. THE THEORETICAL FORMULATION

The ME procedure finds the one pattern which maximizes the probability P of different patterns appearing as the object. The estimated pattern object $\{\hat{Ob}\}$ designated by its set of normalized variables $\{p_i^{(o)}\}$. The variables $\{p_i\}$ are constrained to obey a set of subsidiary conditions.

Near the maximum of the probability $P\{p_i^{(o)}\}$, P behaves quadratically and we may assume it behaves as a multivariate Gaussian distribution. We write symbolically,

$$P\{p_i\} = P\{p_i^{(o)}\} \exp\left\{-\frac{1}{2}\left(\frac{p_i - p_i^{(o)}}{\sigma_i}\right)^2\right\}, \qquad (34)$$

where the $\{\sigma_i\}$ are the standard deviations of the estimate to be calculated. This is illustrated in Figure 15 as though it were a one-dimensional problem.

We will ultimately use the calculated $\{\sigma_i\}$ in determining the confidence of the ME estimated values $\{p_i^{(o)}\}$. A small σ, referring to Figure 15, means that only the values of $\{p\}$ nearly equal to $\{p^{(o)}\}$ have a high probability $P\{p\}$; thus we can rely on the estimate more than if the $\{\sigma\}$ were large. More conventionally, we would normalize the probability to unit integrated volume ("area" in Figure 15) and consider it to be a probability density function. Then at $\{p^{(o)} \pm 1.\sigma\}$, the integrated area (twice the shaded area in Figure 15) would represent

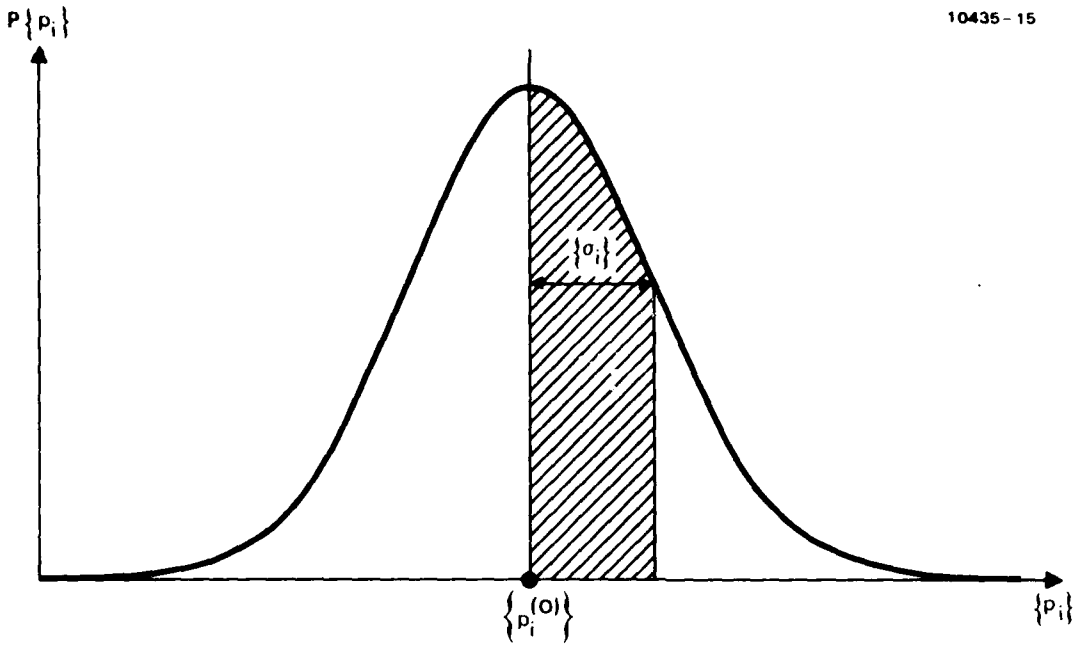


Figure 15. Schematic one-dimensional illustration of the assumed Gaussian behavior of P near its maximum. If P is normalized it becomes the probability density function and the integrated "area" becomes the cumulative probability.

the probability that Ob{p} lies within the range $\{p^{(o)} \pm \cdot\sigma\}$. In one dimension the probability would be 0.68. We could say in that case that the "confidence interval" $p^{(o)} \pm\sigma$ has a "confidence coefficient" of 68% and that the "confidence limits" we set are $\pm 1 \cdot \sigma$.

Our problem differs from the usual statistical problem in several respects. In our problem both the mean, or more correctly, the most probable $\{p^o\}$, are estimated as parameters and not measured as a stochastic variable, and $\{\sigma\}$ are also calculated rather than measured. Furthermore we derive the actual probability distribution function P or density function P_D from the ME theoretical formulation, although we approximate it by a Gaussian near its maximum. We have typically only one (multidimensional) measurement from which we calculate both a

(multidimensional) set of $\{p^o\}$ and one of $\{\sigma\}$ along with the probability distribution of $\{p\}$ as well. We could determine the exact shape of the multivariate function $P\{p\}$ and we would not have to approximate it near the maximum by a multivariate Gaussian, but the computational problem is too large to make that option practical.

The schematic quasi one-dimensional drawing of Figure 15 can be somewhat misleading, so we show in Figure 16 a two-dimensional Gaussian distribution. The two-dimensional example is easily interpreted geometrically and is readily generalizable to the multidimensional case of interest in this problem. In two dimensions,

$$P(p_1,p_2) = P(p_1^{(o)},p_2^{(o)}) \exp \left\{ -\frac{1}{2}\frac{1}{1-\rho^2} \left[\left(\frac{p_1-p_1^{(o)}}{\sigma_1}\right)^2 - 2\rho \frac{p_1-p_1^{(o)}}{\sigma_1} \cdot \frac{p_2-p_2^{(o)}}{\sigma_2} + \left(\frac{p_2-p_2^{(o)}}{\sigma_2}\right)^2 \right] \right\} \tag{35}$$

where ρ is the correlation parameter that characterizes the stochastic dependence between p_1 and p_2. Except when $\rho = 0$, the axes of the contour ellipses are not parallel to the coordinate system p_1, p_2, and the length of the axes of the ellipses depend on both $\{\sigma\}$ and ρ. Equation 35 can be normalized so that its two-dimensional integral is unity to make a probability-density function:

$$P_D(p_1,p_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \{\text{as in (35)}\} . \tag{36}$$

Curves of equal probability on planes parallel to the p_1p_2 plane are ellipses on the P surface of Figure 16, as can be seen from the exponent of Eq. 35. These ellipses may be projected onto the p_1,p_2 plane as shown in Figure 17.

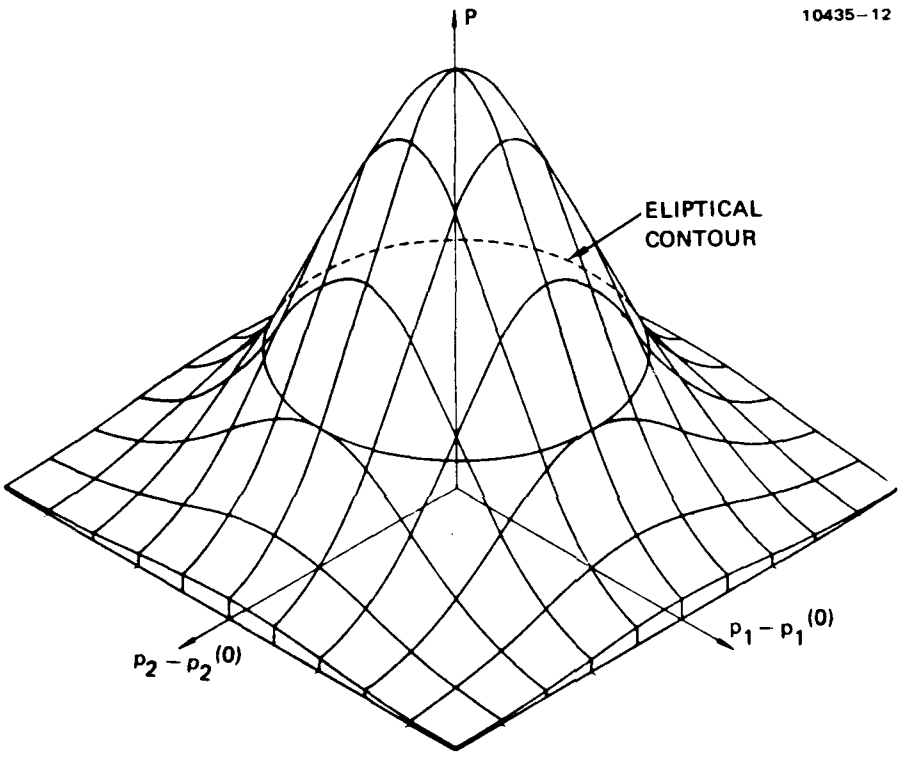


Figure 16. Distribution surface for a two-dimensional normal distribution.

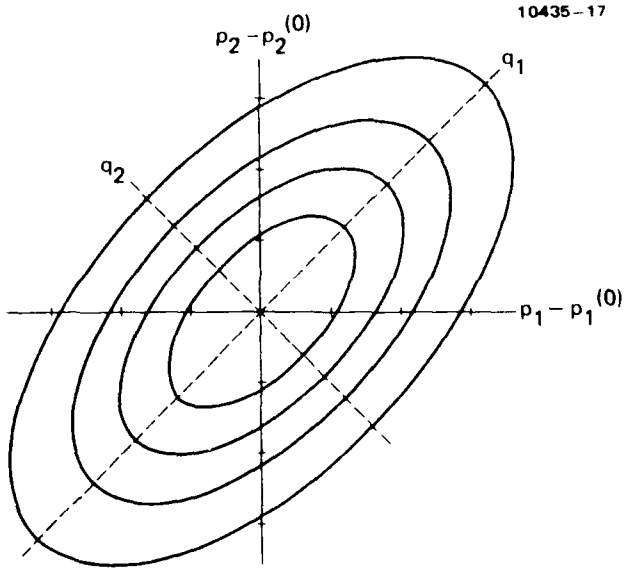


Figure 17. Concentration ellipses of a bivariate normal density. These are contours of equal probability P projected on the $p_1 p_2$ plane.

The contour ellipses corresponding to Eq. 36 are of the form

$$\frac{1}{1-\rho^2}\left[\left(\frac{p_1-p_1^{(o)}}{\sigma_1}\right)^2 - 2\rho\,\frac{p_1-p_1^{(o)}}{\sigma_1}\cdot\frac{p_2-p_2^{(o)}}{\sigma_2} + \left(\frac{p_2^{(o)}-p_2}{\sigma_2}\right)^2\right] = C^2, \qquad (37)$$

where C is a constant. It is always possible to make a linear orthogonal transformation of the variables to a new set of variables which are stochastically independent and normally distributed. In effect, it is a rigid rotation of the coordinates to a new system where the axes of the ellipse lie along the new coordinate axes q_i (see Figure 17). With the new correlation coefficient equal to zero, the new variances σ_q are proportional to the axis lengths of the ellipse.

The new equation of (the same) density function P_D (36) can then be written

$$P_D(q_1,q_2) = \frac{1}{2\pi\sigma_{q_1}\sigma_{q_2}} \exp\left[-\frac{1}{2}\left(\frac{q_1^2}{\sigma_{q_1}^2} + \frac{q_2^2}{\sigma_{q_2}^2}\right)\right], \qquad (38)$$

and the contour ellipses corresponding to those of Eq. 37 can be written

$$\frac{q_1^2}{\sigma_{q_1}^2} + \frac{q_2^2}{\sigma_{q_2}^2} = C^2 \quad . \qquad (39)$$

This form is the sum of squares of two stochastically independent variables and, therefore, has a χ^2 distribution with two degrees of freedom. The probability that (q_1,q_2) is inside the ellipse (Eq. 39)

given by C^2, is therefore equal to the P-fractal of the cumulative χ_P^2 distribution function for two degrees of freedom f,

$$\frac{q_1^2}{\sigma_{q_1}^2} + \frac{q_2^2}{\sigma_{q_2}^2} = \chi_P^2 \ (f=2) \quad . \tag{40}$$

Equation 39 is exactly the same ellipse as before (Eq. 37), which was expressed in terms of the correlated variables p. The only purpose of the transformation here is to show the equivalence as a sum of squares of independent variables. But there are other uses of the transformation which we will describe below.

For three variables, (p_1, p_2, p_3), the trivariate Gaussian distribution p cannot be drawn, but we show in Figure 18 one three-dimensional concentration ellipsoid represented in its rotated coordinates $\{q_i\}$. The lengths of the three principal axes $\{\sigma_i\}$ are shown as well. The probability that a point is inside this particular ellipsoid is given by the P-fractal of the cumulative χ_P^2 distribution function for f = 3 degrees of freedom. We have chosen $C^2 = \chi_P^2 = 1$ for Figure 18. Tables of χ^2 show that the confidence level P here is 20%.

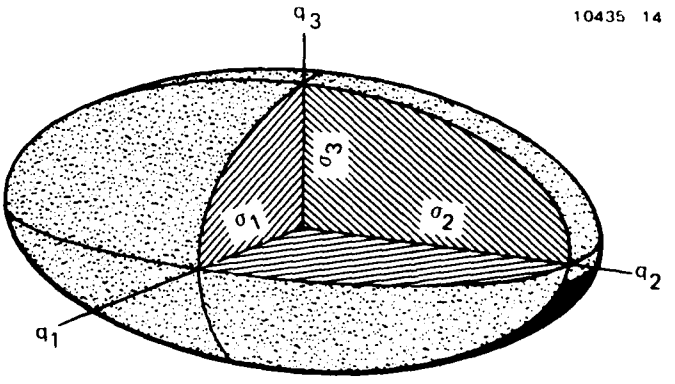


Figure 18. A concentration ellipsoid for the trivariate Gaussian probability distribution.

In one sense, the relation (Eq. 40) or its multidimensional equivalent for larger number of degrees of freedom, f, is a solution of the problem of finding the confidence region for the ME estimates. The $\{\sigma_q\}$ and $\{q\}$ are calculated (or equivalently the $\{\rho\}$ and $\{\sigma\}$) for the object estimate $\{p^{(o)}\}$ and, for a chosen confidence level, say P = 95%, one can compute whether any given point $\{p\}$, which represents a multidimensional estimate $Ob\{p\}$, lies within the hyperellipsoid given by P. But a graphical or pictorial representation of this condition, even in two-dimensional cases, is already problematical.

Consider a two-dimensional ME estimate where only two parameters are being estimated, e.g., p_1 and p_2. Suppose they are to be represented graphically as in Figure 19(a) as the strengths of the vector components (or the function) $\{p_i\}$. (Perhaps this would seem a more reasonable procedure if we had a larger number than merely two.) In Figure 19(a) suppose we attempt to show by two individual "error bars" or confidence intervals, a confidence band attempting to represent one of the accumulation ellipses of Figure 17, say for the 95% confidence level P. This is replotted in Figure 19(b) around $(p_1^{(o)}, p_2^{(o)})$. In 19(b) objects $\{p_i\}$ are described by the coordinates p_i or the vector $\{p_i\}$ in two-dimensional space. Imagine a point going around the circumference of that ellipse, its p_1 and p_2 values going from maximum to minimum separately, but in a concerted fashion. If we label the maximum values of p_1 by 1 and p_2 by 2 in Figure 19(b), we can replot these points on the confidence intervals in Figure 19(a). Notice that points outside the ellipse, such as points 4 or 5, are still represented by points within these confidence intervals which were determined by the maximum extent of the ellipse! By admitting points such as 4 or 5 we have seemingly enlarged the confidence region in a distorted way to include some portion of fractal regions, higher than our chosen 95% value. Instead of the desired correct ellipse, we have represented a circumscribed rectangle. There is no precise way to represent the ellipse or hyperellipsoid in the fashion of Figures 19(a) or 19(c) as confidence bands. An inscribed rectangle for example, instead of a circumscribed one, would indeed decrease, or

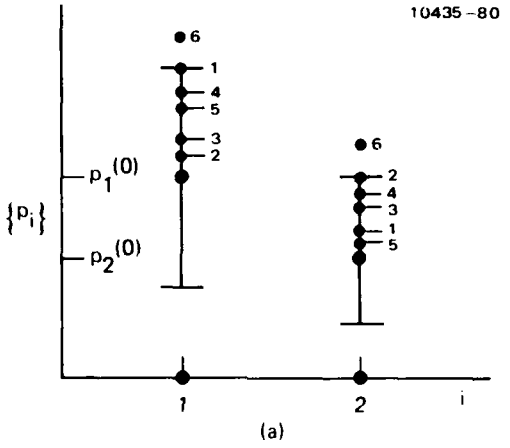


Figure 19(a). Graphical representation of the two-dimensional, i.e., two parameter, estimates $\{p_1^{(o)}, p_2^{(o)}\}$ including confidence intervals to make a confidence band. Here objects are described by plotting the amplitude of the components of $\{p_i\}$ considered as a vector or a function.

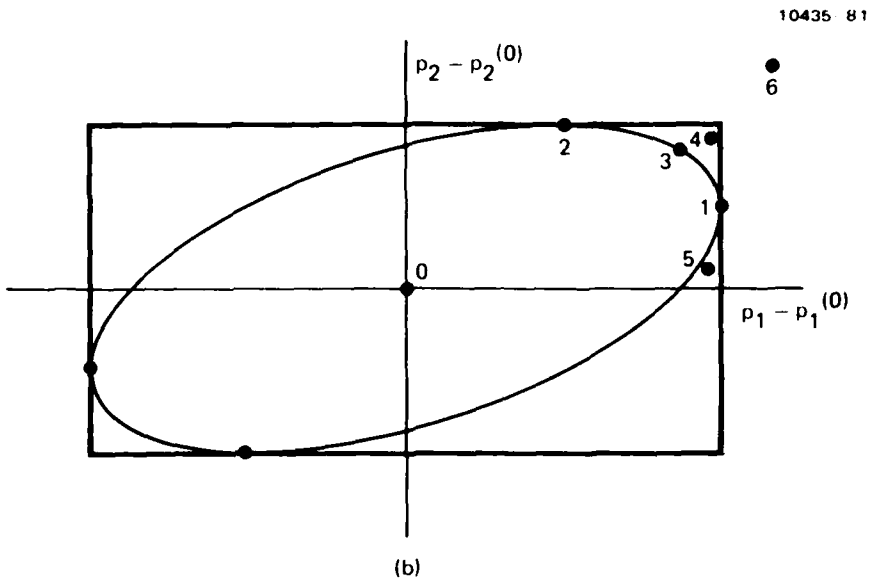


Figure 19(b). The accumulation ellipse and circumscribed rectangle from which 19(a) is derived. Here objects are described by points in multi-dimensional (here two-dimensional) space with coordinates or vector position $\{p_i\}$.

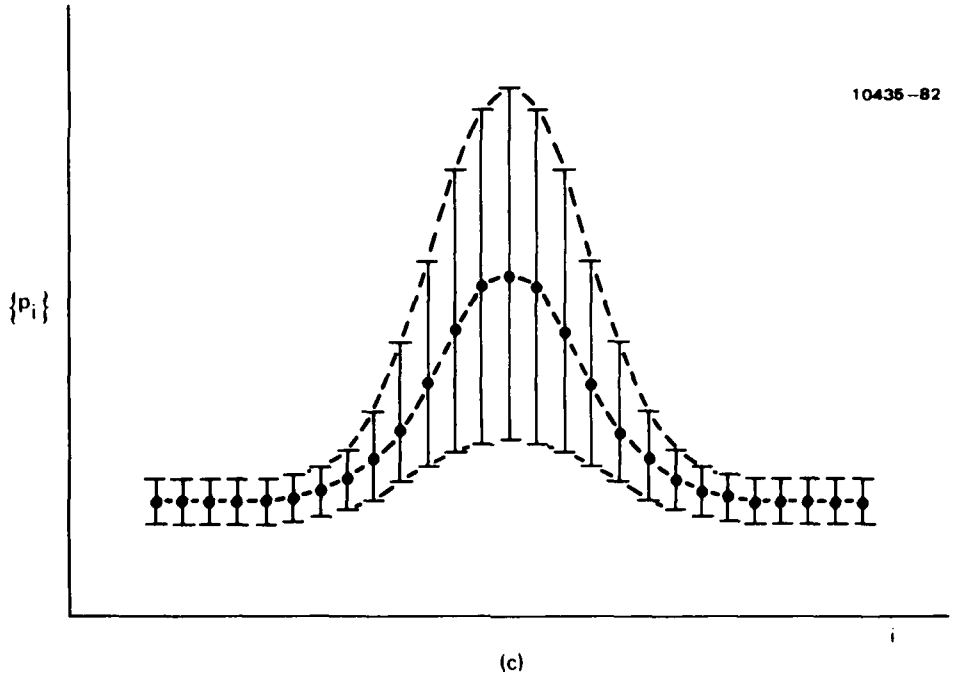


Figure 19(c). Graphical representation of a multidimensional estimate $\{p_i\}$, including confidence band, as a generalization of 19(a).

perhaps even eliminate improperly included extraneous regions, but it would cut out some of the ellipse as well.

We choose a graphical method which will at least represent the ellipse or hyperellipsoid exactly. Yet it does not always provide visual understanding of whether a given object estimate point in multi-dimensional P space with coordinates $\{p_i\}$, lies inside or outside a particular accumulation hyperellipsoid by merely displaying confidence bands laid out as one-dimensional graphs of the strength of the components of the vector $\{p_i^{(o)}\}$ + {confidence interval} (except in some interesting special cases).

The method we choose, following our present geometrical interpretation, is to use the endpoints of the ellipse or hyperellipsoid, i.e., the lengths along the principal axes, rather than a circumscribed rectangle as in the example of Figure 19(b). For any desired ellipse, these lengths, which are proportional to the variances σ_q calculated in rotated, uncorrelated q space, are then linearly projected onto p space as components, to be used as confidence intervals. In Figure 20(a) we show such a projection for the two-dimensional case.

The endpoint q_1 along the largest axis of largest σ_{q_1} is projected as $[p_1-p_2^{(o)}]^{(1)}$ along the p_1 axis and as $[p_2-p_2^{(o)}]^{(1)}$ along the p_2 axis. Similarly, the endpoint q_2 of the smallest axis of σ_{q_2} is projected as $[p_1-p_1^{(o)}]^{(2)}$ and $[p_2-p_2^{(o)}]^{(2)}$. In Figure 20(b) we plot the amplitudes of the projected components as confidence intervals separately for the largest and for the smallest axes. Note that, unlike the method of Figure 19, we now have two separate confidence bands, one associated with each of the two principal axes. In Figure 20(c) we show the generalization to the multidimensional case of N estimates or variables $\{p_i\}$, $1 \le i < N$. There are a set N different confidence bands. Often only one or only the first few are of any significant magnitude, thereby greatly simplifying the problem of interpretation.

The scheme shown in Figure 20(c) is the one we adopt in this report for our 49-dimensional study example. The geometric interpretation discussed above has a well-known algebraic counterpart in the

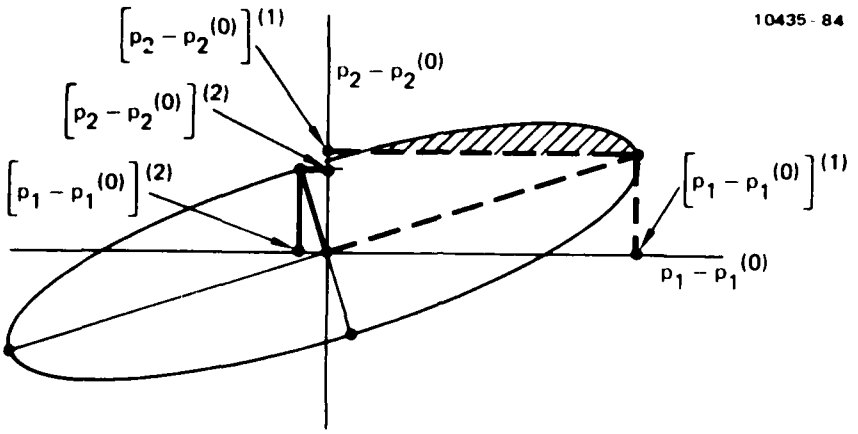


Figure 20(a). The accumulation ellipse of the two-dimensional i.e., two-parameter estimates $\{p_1^{(o)}, p_2^{(o)}\}$ with projections of the endpoints of the major and minor axes onto the p coordinate axes.

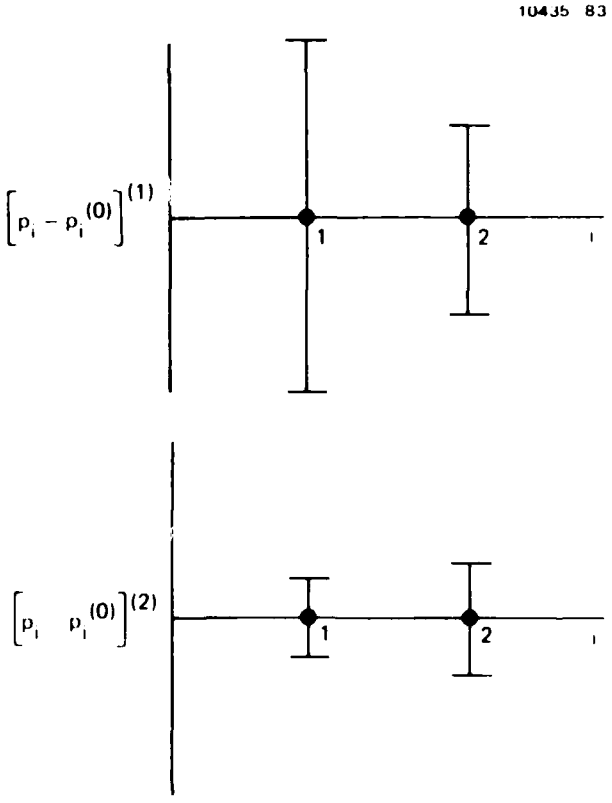


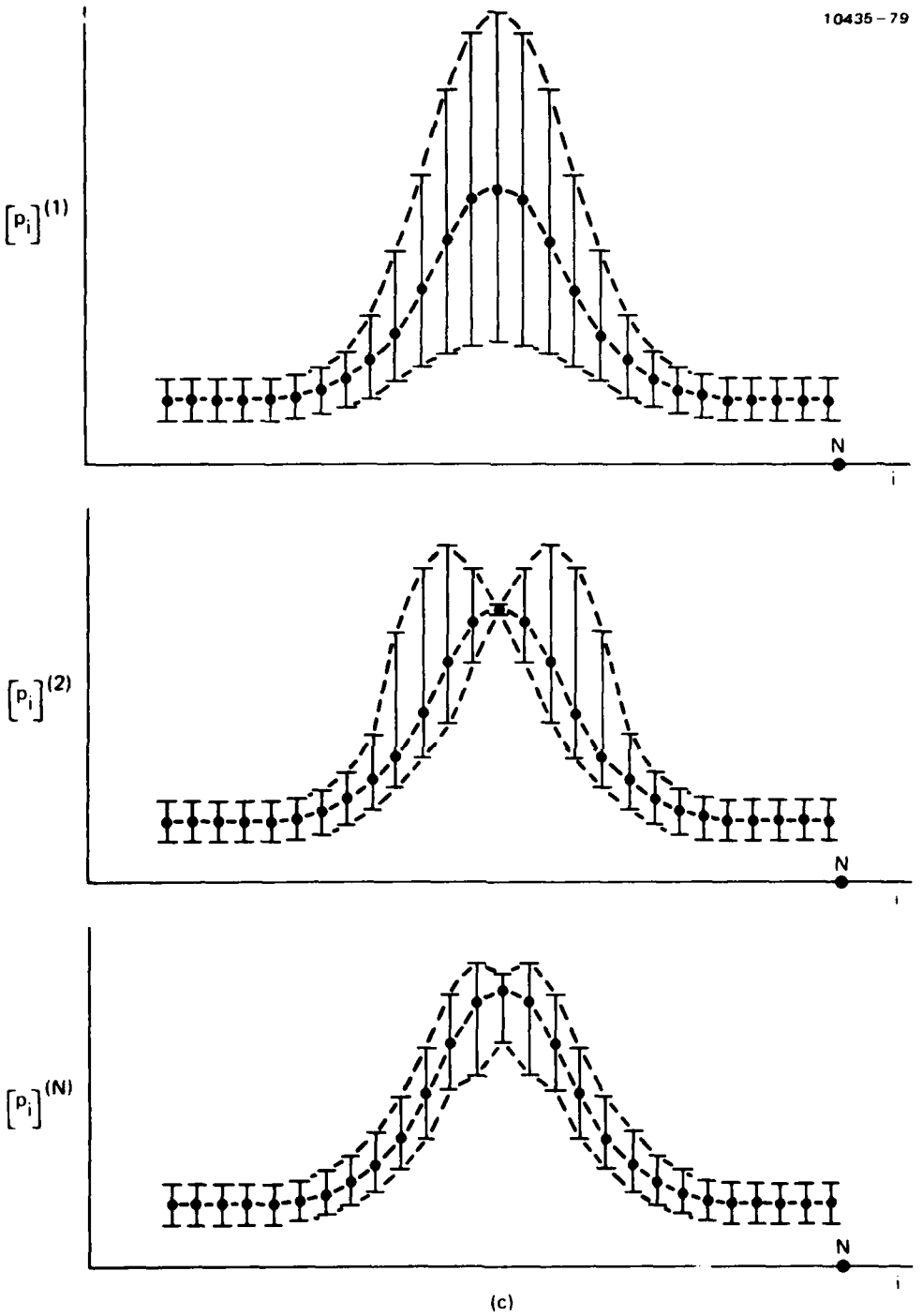
Figure 20(b). The estimates $\{p_1^{(o)}, p_2^{(o)}\}$ including projected confidence intervals to make confidence bands. A separate plot is given for each principal axis of the ellipse of 20(a).

determination of eigenvectors (or characteristic vectors) and the associated eigenvalues (or characteristic values). The directions of the principal axes of the concentration ellipsoids are specified by direction cosines which are given by the components of the normalized eigenvectors of a generalized variance or covariance matrix that includes off-diagonal terms to account for correlations. Each eigenvector belongs to an eigenvalue, and the length of any principal axis, for a particular concentration ellipsoid, is proportional to the square root of its associated eigenvalue.

It can be shown that the eigenvectors are orthogonal; therefore, this method of principal-axis transformation results in uncorrelated variates whose variances are proportional to the axis lengths of any specific concentration hyperellipsoid. To display the confidence limits on the original unrotated object estimate, we project these axis lengths back to the original coordinate axes (p) as illustrated in Figure 20(c).

B. EXAMPLES

In our ME confidence examples we choose for measurement data constraints, images in Fourier space derived from a standard test object. That object consists of two spikes symmetrically disposed on a support of $1 \leq i \leq 49$ at positions i=24 and i=26. The "images" in Fourier space differ only in the aperture or bandwidth accepted for their measurements. The scaled apertures were varied from an almost full aperture of 48 units, out of 49, giving almost enough resolution to resolve the peaks, through 24, 12 and 6 units with approximately an eight-fold maximum degration in resolution. This afforded opportunities for as much as an eight-fold possible enhancement by ME superresolution. The gamut of inverse noise or fluctuation temperature β was $2^{0} \leq \beta \leq 2^{19}$. Larger values of β did not converge. From all these examples studied, we will only present ten particular ones in this



(c)

Figure 20(c). The multidimensional estimates, including the set of confidence bands, one from each principal axis of the hyperellipsoid, as a generalization of 20(b).

report, chosen to illustrate the main results. A listing is given in Table 1. The ME calculations were done with the log B formulation in the examples. This formulation used in these examples, it will be recalled, is appropriate for signal statistics that are Gaussian in complex amplitude or exponential in intensity.

Table 1. Parameters for Computations

Aperture Size	Potential Factor of Superresolution	Inverse Temperatures, $\log_2 \beta$
6	8	0, 18
12	4	0, 8, 17
24	2	0, 16, 19
48	1	0, 18

The results are arranged in sets according to the aperture sizes shown in Table 1. In each set two or more values of β are chosen for examination. Heading each set there is a display of the ME estimates for the given aperture, one estimate for every value of β in the gamut of β. The hidden line graphic displays sometimes, for clarity, require a separate picture for the positive and the negative values of the function displayed. It should be noted that the discrete points of the functions are connected by straight lines often giving the functions the appearance of a set of triangular spikes. This is merely an artifact of the representation.

The ME estimates for aperture size 6 and $2^0 \leq \beta \leq 2^{18}$ are shown in Figure 21. The estimated objects, $Ob\{p_i\}$ are plotted vertically versus i, with the estimates for various values of β placed one behind the other in a third dimension. The base of the figure is zero. For this small aperture size, superresolution was not achieved, even at the coldest temperature that we could simulate corresponding to $\beta = 2^{18}$.

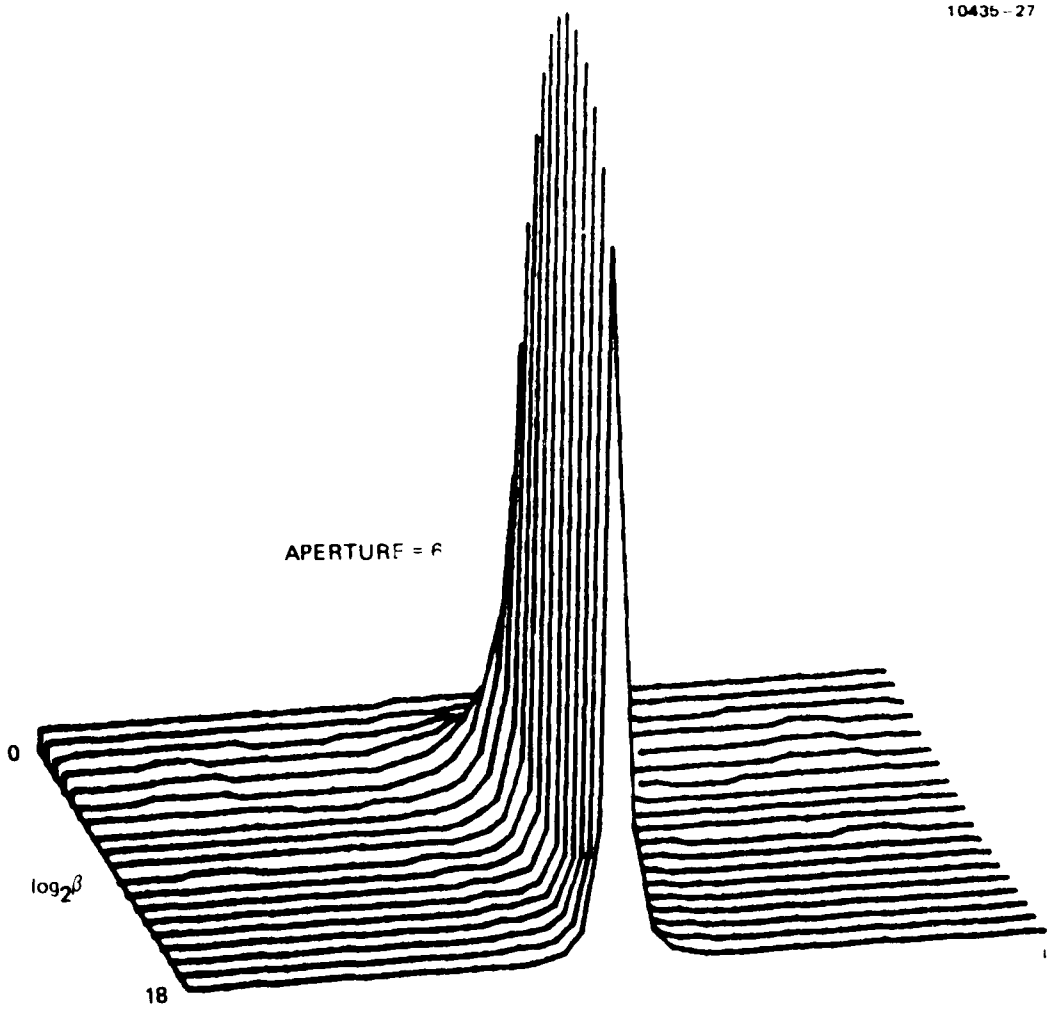


Figure 21. The ME estimates $\{p_i^{(o)} : 1 \le i \le 49\}$, for aperture size 6 and for $2^0 \le \beta \le 2^{18}$.

The two peaks at i = 24, i = 26 are seen to be rising, however, along with the diminishing of the central peak for γ values greater than 2^{13}. These, and all the other ME estimates we exhibit in this report, have been constrained to be normalized to unit energy, power or area, so there are only 48 "independent" variables, which are further constrained in the ME method (as explained previously). This normalization explains the general rise in the background with the fall in the peak as the estimate tends to uniformity when γ approaches zero.

The orthonormal eigenvectors of the confidence hyperellipsoid for $\gamma = 2^{0}$ and aperture = 6 are shown in Figure 22(a) and 22(b). Since these q-space vectors are all still normalized, in the figure they represent a hypersphere. The method used is to plot the amplitude of the p-space components of the eigenvector (or eigenfunction) as a continuous curve, after the procedure illustrated in Figure 20 for two dimensions. The eigenvectors are rank ordered according to the strength of their associated eigenvalues, the greatest called the first and the last called the 49th. The hidden line plot hides all but the first of the negative peaks so we separately plot the positive peaks up in Figure 22(a) and the negative peaks up in Figure 22(b). The approximate zero level is halfway between the base and peaks. These eigenvectors were then normalized according to the square roots of their respective eigenvalues to form $\{\ \}_i$ and scaled by $\sqrt{\gamma^2}$. The result for 49 dimensions, $\gamma = 2^{0}$, and aperture of 6 is shown in Figure 22(c). The base of the figure represents minus one and the top mesa represents zero. The vectors composed of the sum of the ME estimates $\{p_i^{(o)}\}$ plus $\sqrt{\gamma^2} \cdot \{\ \}_i$ are shown in Figure 22(d) for our present case of aperture size 6 and $\gamma = 2^{0}$. The case $\{p_i^{(o)}\}$ minus $\sqrt{\gamma^2} \cdot \{\ \}_i$ is shown in Figure 22(e). The zero of $\{p_i\}$ is the base of Figure 22(d) and 22(e). Figures 22(d) and 22(e) are the hidden-line versions of the separate upper and lower confidence limits shown in the earlier illustrative example given in Figure 20(c). Here $\sqrt{\gamma^2} = 1.05$; larger values would make p negative.

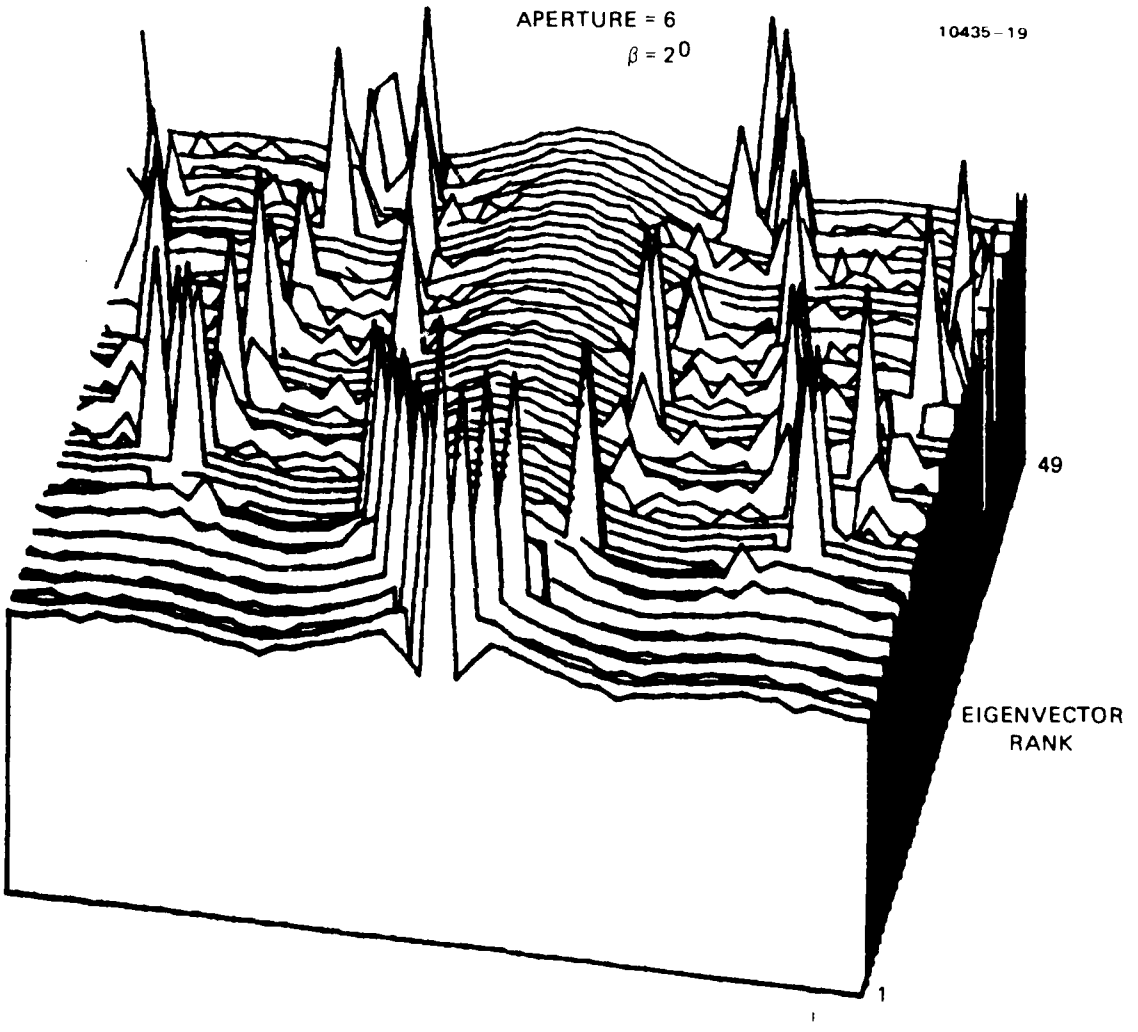


Figure 22(a). Ortho-normal eigenvectors projected on p-space for aperture size 6 and $\varepsilon = 2^0$. Positive peaks.

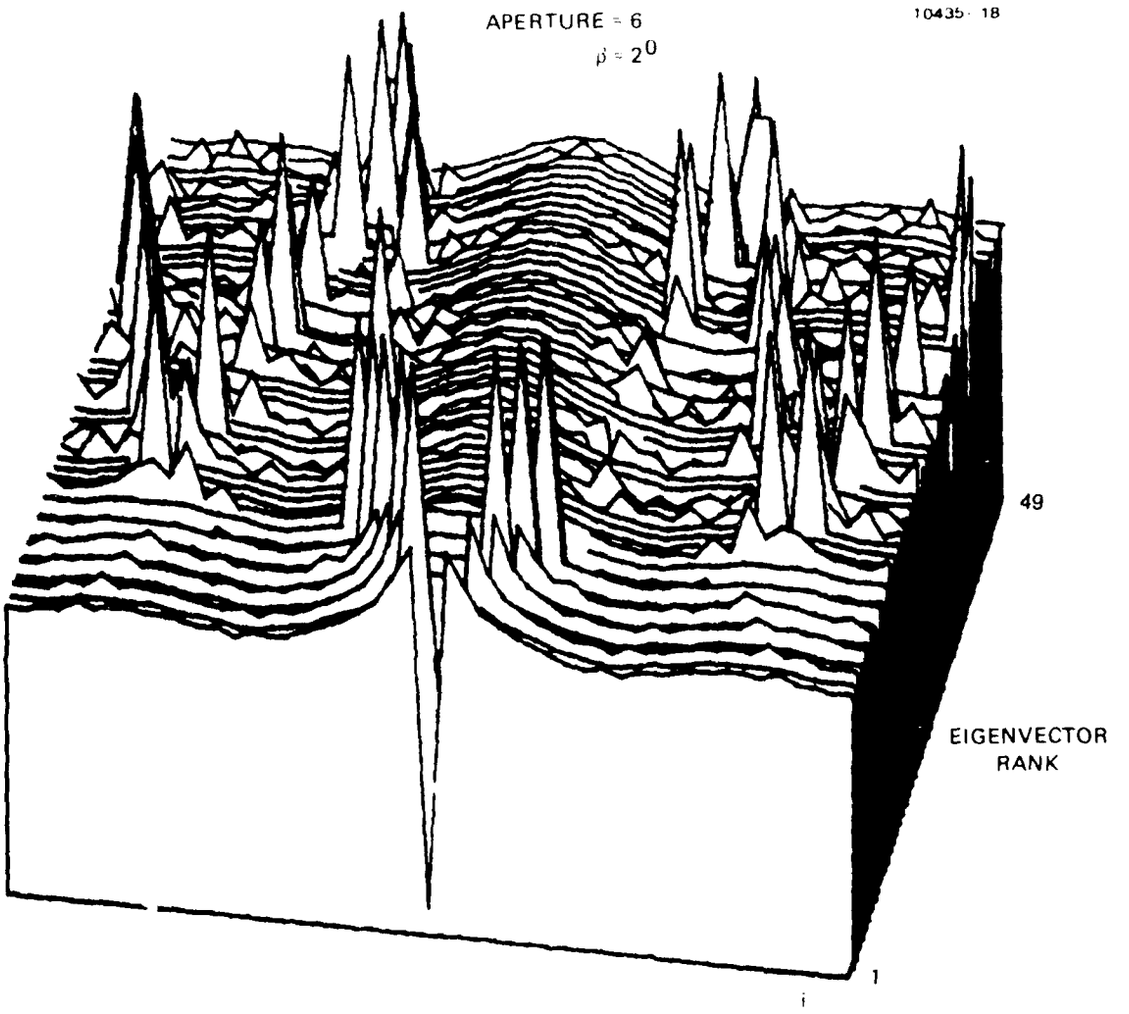


Figure 22(b). Ortho-normal eigenvectors projected on p-r space for aperture size 6 and $\rho = 2^0$. Negative peaks.

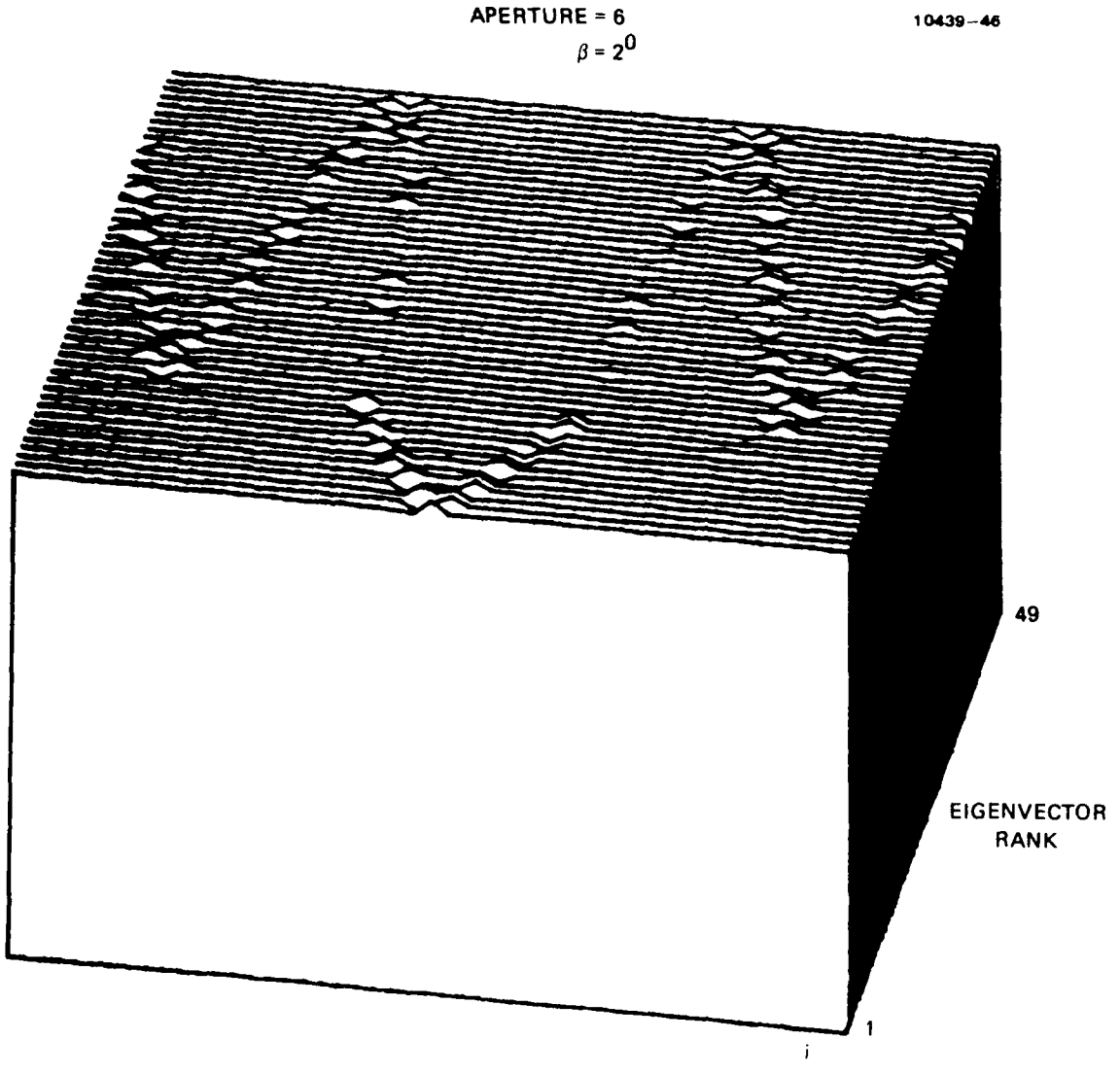


Figure 22(c). Eigenvectors rescaled and projected to p-space as $1.05\ \{\sigma_i\}$ for aperture 6, $\beta = 2°$.

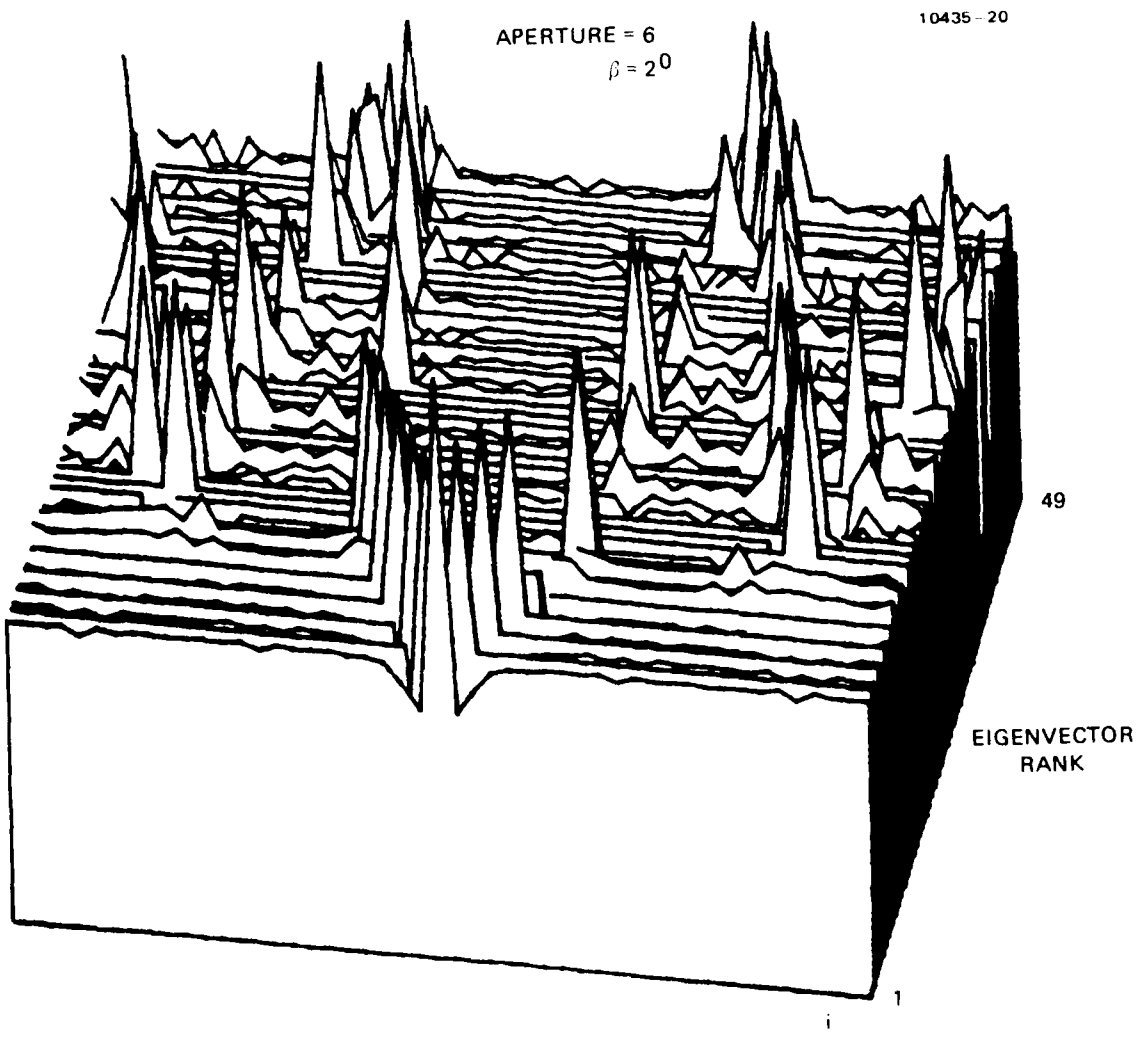


Figure 22(d). The ME estimates $\{p_i^{(o)}\}$ plus 1.05 $\{\sigma_i\}$ for aperture 6, $\beta = 2°$.

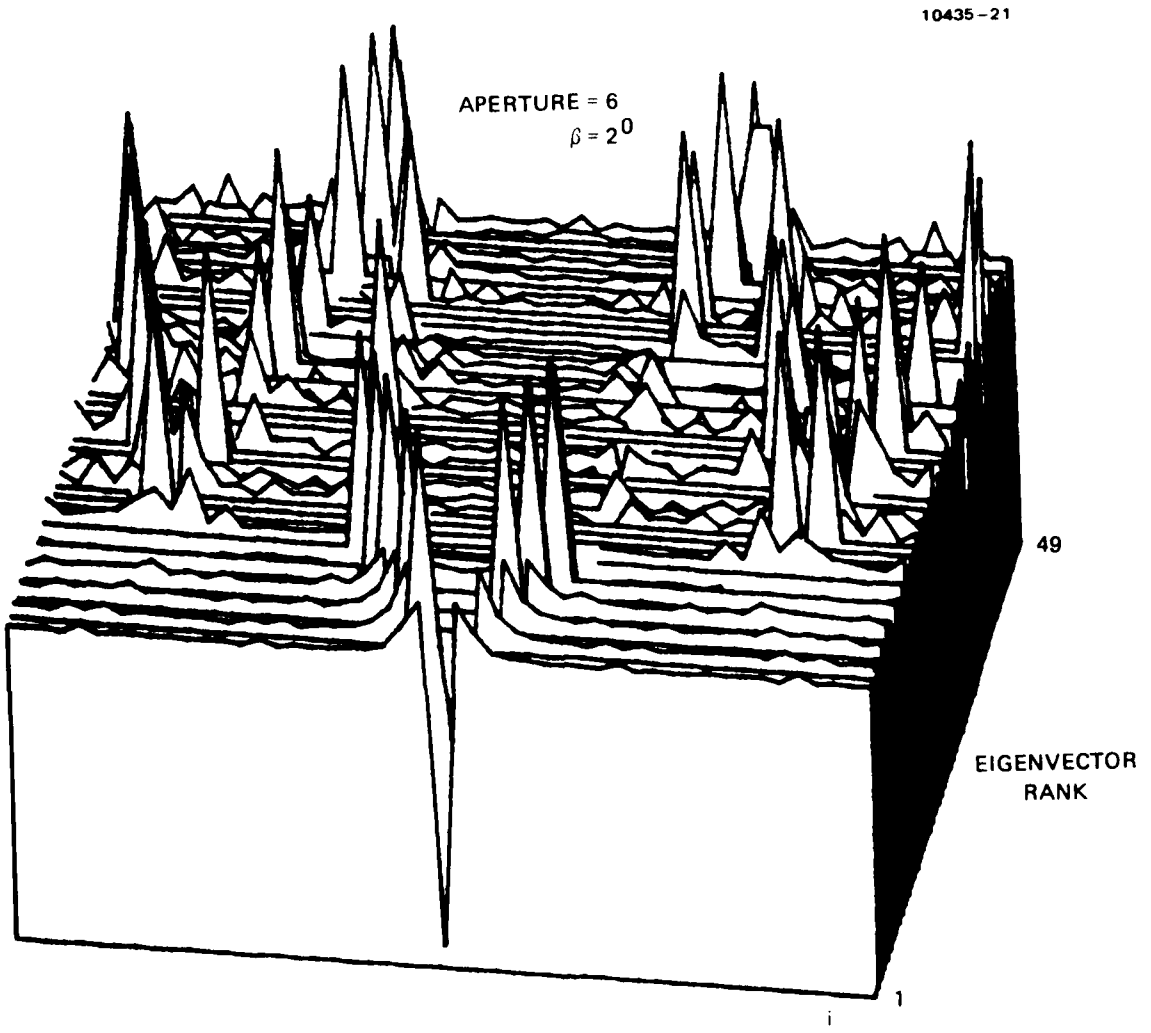


Figure 22(e). The ME estimate $\{p_i^{(o)}\}$ minus 1.05 $\{\nu_i\}$ for aperture 6, $\beta = 2°$.

The real probability density function P_D for our problem would not have any negative values of $\{p_i\}$, but our multivariate Gaussian approximation has tails to infinity. For a large enough χ^2, p_i could become negative. We note in Table 2 that the maximum value of $\sqrt{\chi^2}$ to avoid this was found to vary from slightly over 1 to $\sqrt{2}$ as r varies from small to large values. These are the values of $\sqrt{\chi^2}$ used in the examples plotted here. Interpreted as χ^2 of 48 degrees of freedom, these confidence limits would represent extremely small confidence coefficients. Interpreted as for one degree of freedom, for the large r cases where all but one σ_1 are essentially zero, they represent confidence coefficients of 70% to 84%. A possible explanation for this upper band of the range of multiplying factors, $1 \leq \sqrt{\chi^2} \leq \sqrt{2}$ for our multivariate Gaussian approximate probability density function may lie in the fact that we are dealing with the case of Gaussian statistics. Our "log p" formulation used in these examples is appropriate for this statistic. In the Gaussian case, $\overline{\Delta p^2} = (\overline{p})^2$, but recall the definition

$$\sigma \equiv \sqrt{\overline{\Delta p^2}} \quad .$$

Thus for Gaussian statistics, $\sigma = \overline{p}$. Therefore in this case, $p = \overline{p} - 1 \cdot \sigma$ is already equal to zero. Any greater multiplier of σ will make p negative. This one-dimensional example may point the way to a proper understanding of our empirical observation.

We see in Figures 22(c), (d), and (e) that the hyperellipsoid accumulation is almost hyperspherical. There is no one σ that is remarkably greater than the others. The confidence region surrounds the entire estimate almost uniformly. This will prove to be true for all cases of small r.

Recalling that the σ_i are equal to the square root of the eigenvalues, we see in Table 2 that the cases of $r = 2^0$, for any aperture size, all have an almost flat distribution of σ_i, varying by, at most,

Table 2. Data for Computations

TABLE 2. Data for computations

Aperture	Inverse Temperatures $\log_2 \beta$	Eigenvalues σ_i^2 for $1 \le i \le 49$	σ_1^2/σ_2^2	σ_1^2/σ_{49}^2	Maximum χ^2 for $p_i > 0$
6	0	2.6×10^{-4}, ... 1.9×10^{-4}		1.4	1.05
6	18	8×10^{-3}, 4×10^{-4}, ... 5×10^{-9}	2×10^{1}	2×10^{6}	$\sqrt{2}$
12	0	3.2×10^{-4} ... 1.8×10^{-4}		1.8	1.04
12	8	2.1×10^{-3}, 1.7×10^{-4}, ... 3.8×10^{-6}	1×10^{1}	6×10^{2}	$\sqrt{2}$
12	17	9×10^{-2}, 5×10^{-5}, ... 7×10^{-9}	2×10^{3}	1×10^{7}	$\sqrt{2}$
24	0	3.9×10^{-4}, ... 1.8×10^{-4}		2.2	1.04
24	6	2.8×10^{-2}, 3.3×10^{-4}, ... 8.8×10^{-6}	8×10^{1}	3×10^{3}	$\sqrt{2}$
24	19	1×10^{-1}, 7×10^{-8}, ... 1×10^{-9}	1×10^{6}	1×10^{8}	$\sqrt{2}$
48	0	6.6×10^{-6}, ... 1.6×10^{-4}		4.1	1.03
48	18	1×10^{-1}, 2×10^{-8}, ... 6×10^{-10}	5×10^{6}	2×10^{8}	$\sqrt{2}$

a factor of 2. This can be contrasted to other cases in the table for larger μ values where only one σ_2 is significant and the others are orders of magnitude smaller.

The next example, still with aperture of 6, has $\mu = 2^{18}$. The eigenvectors are shown in Figures 23(a) and (b). For this large μ there are really only two significant σ_i with the second more than four times smaller than the first. This is evident in Figure 23, a plot of the projection of $\sqrt{2} \cdot \sigma_i$. The contribution to the confidence band is only near two unresolved peaks. The ME estimates $\sqrt{2} \cdot \sigma_i$ are shown in Figures 23(d) and (e). Again, only the first two eigenvectors are active and only on the two shoulders of the two unresolved peaks.

Increasing the aperture size to 12 we see in Figure 24 that for $\mu > 2^8$, the ME estimates begin to resolve the two peaks; at $\mu = 2^{18}$ they are far more resolved than demanded by the Rayleigh criterion. This is a four-fold increase in the resolution of this aperture.

For this aperture of size 12 and $\mu = 2^0$ the series of plots, Figures 25(a) through (f) describe the nature of the confidence bands. The peaks are not resolved (Figure 25(f)), and the base represents $p_i = 0$). The confidence bands, again as in the previous $\mu = 2^0$ case, are derived from a hypersphere as all the σ_i are closely equal (see Table 2) the smallest being 3/4 of the largest.

For $\mu = 2^8$ and aperture 12, the ratio of the first two σ_i are $\sigma_2/\sigma_1 = 0.28$. Thus we must consider both. Figure 26(c) shows the relative strengths of their associated eigenvectors. The base of Figure 26(c) represents minus one. The effect on the confidence in the estimates of p_i can be seen in the first two rows of Figures 26(d) and (e) as a broadening of the unresolved single spike at the positions $i = 24$ and $i = 26$, where the two object spikes might be expected to occur.

For $\mu = 2^{17}$ and aperture 12, the ratio of the first two σ_i is $\sigma_2/\sigma_1 = 2^{-2}$, so only the first has an important influence. That only

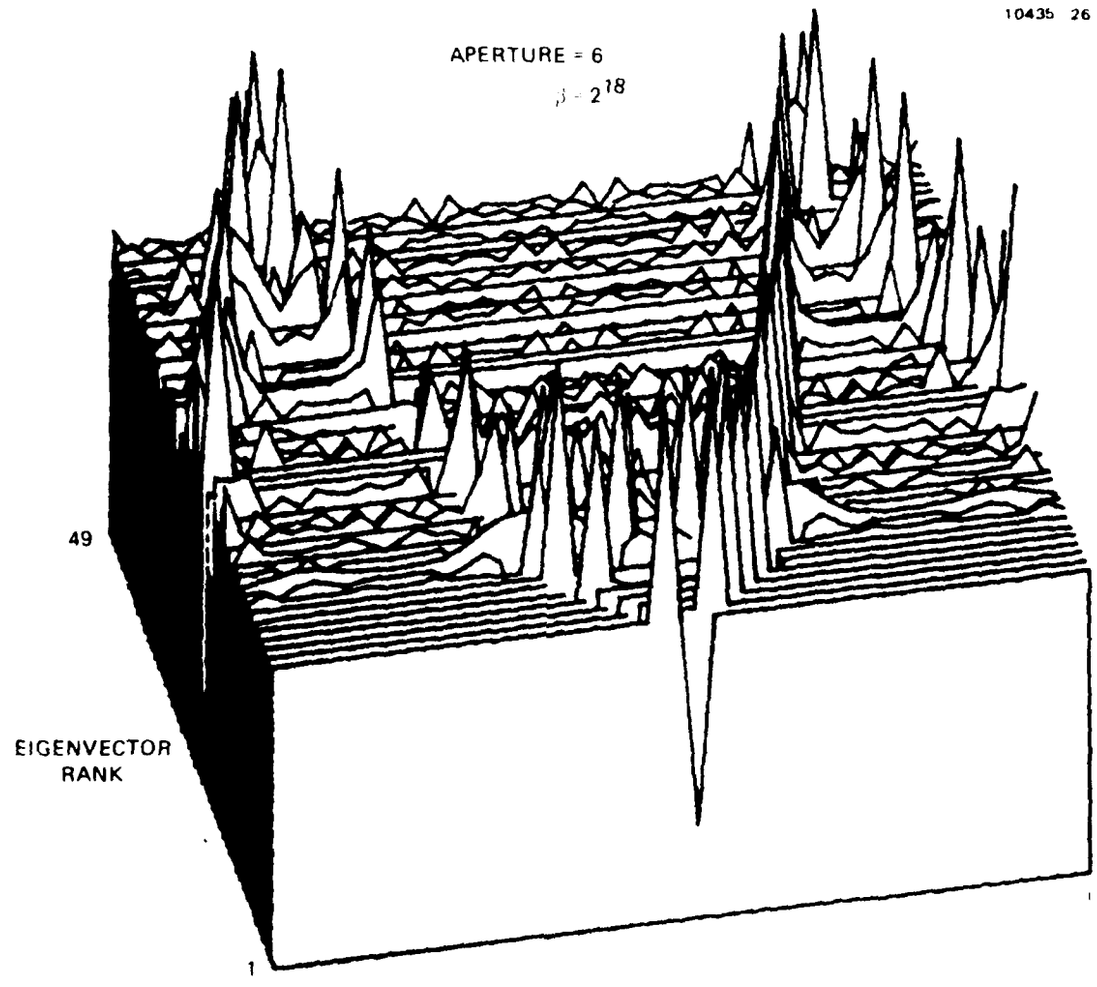


Figure 23(a). Ortho-normal eigenvectors projected on p-space for aperture size 6 and $\beta = 2^{18}$. Positive peaks.

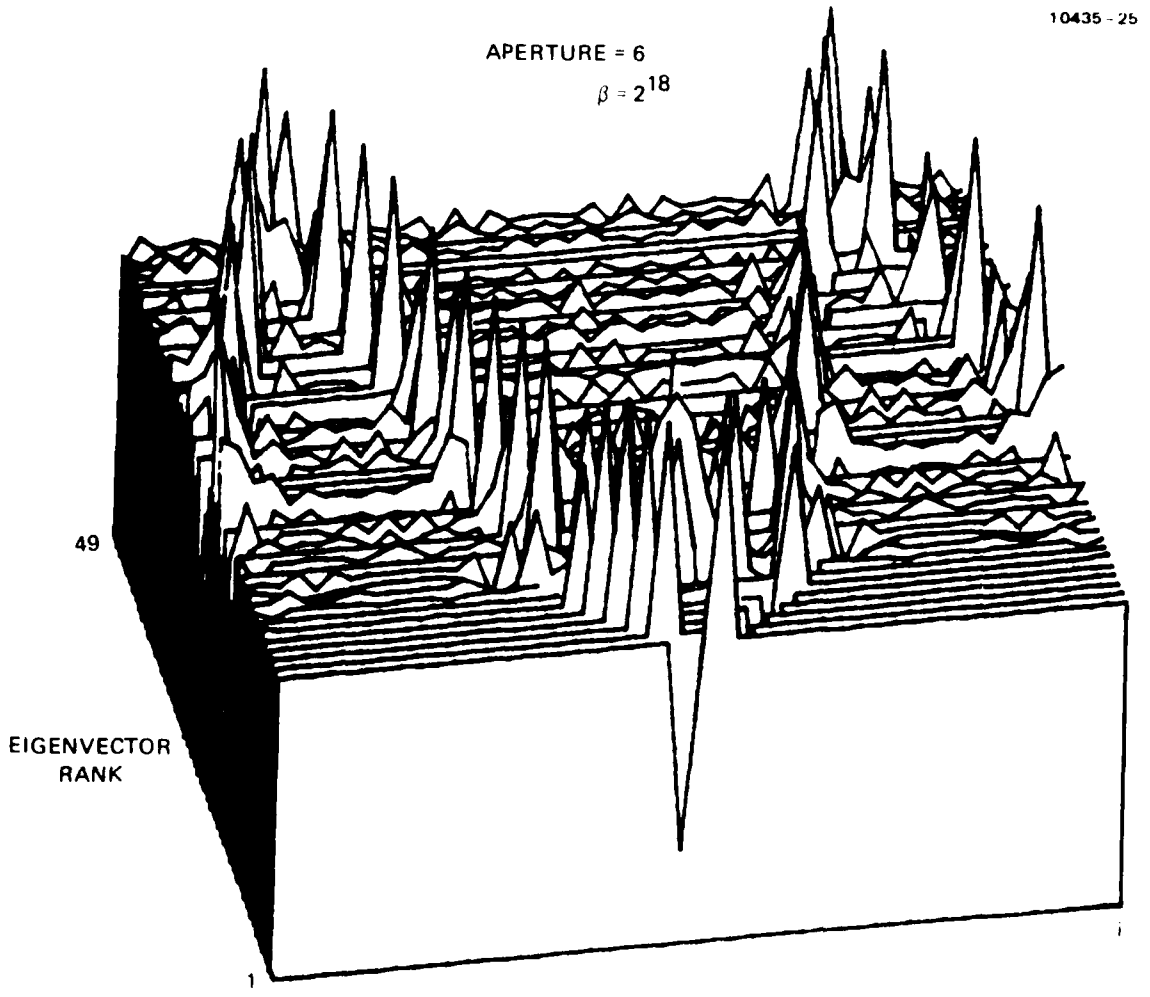


Figure 23(b). Ortho-normal eigenvectors projected on p-space for aperture size 6 and $\beta = 2^{18}$. Negative peaks.

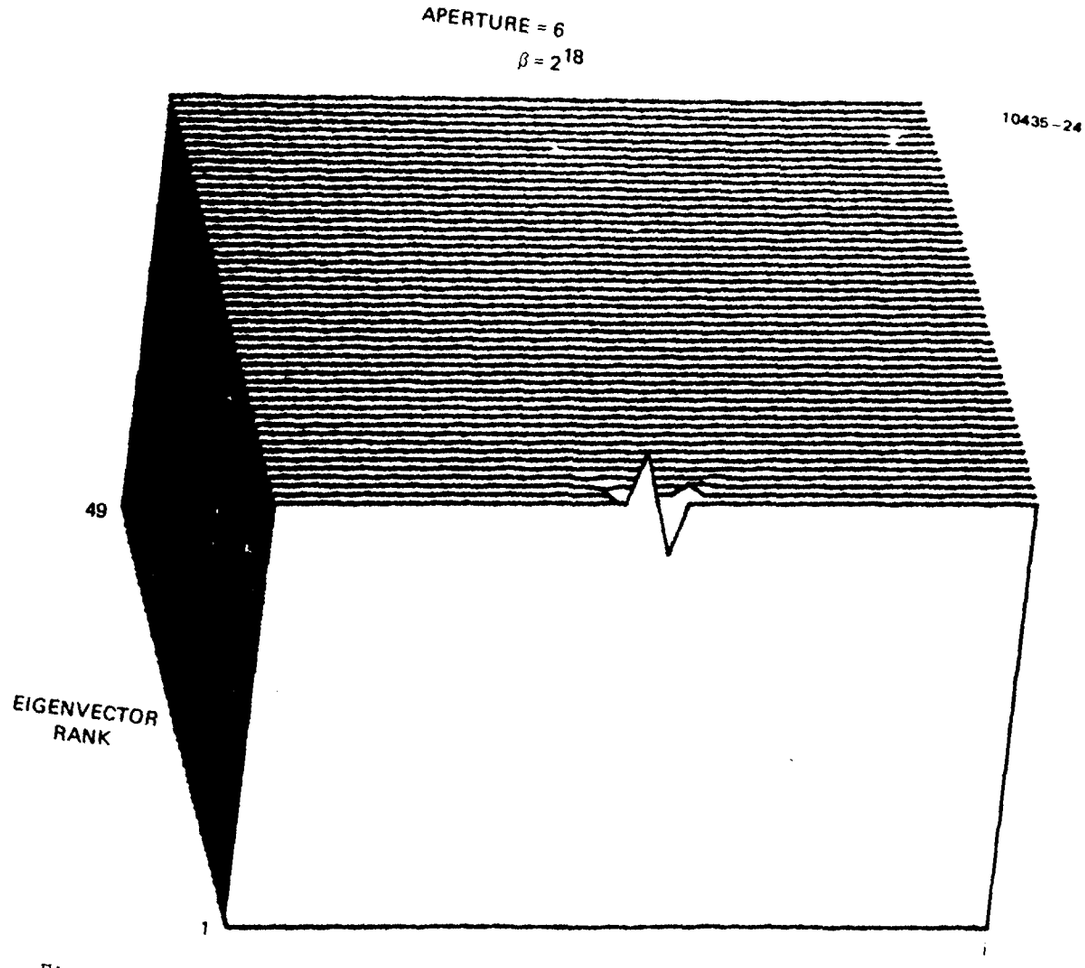
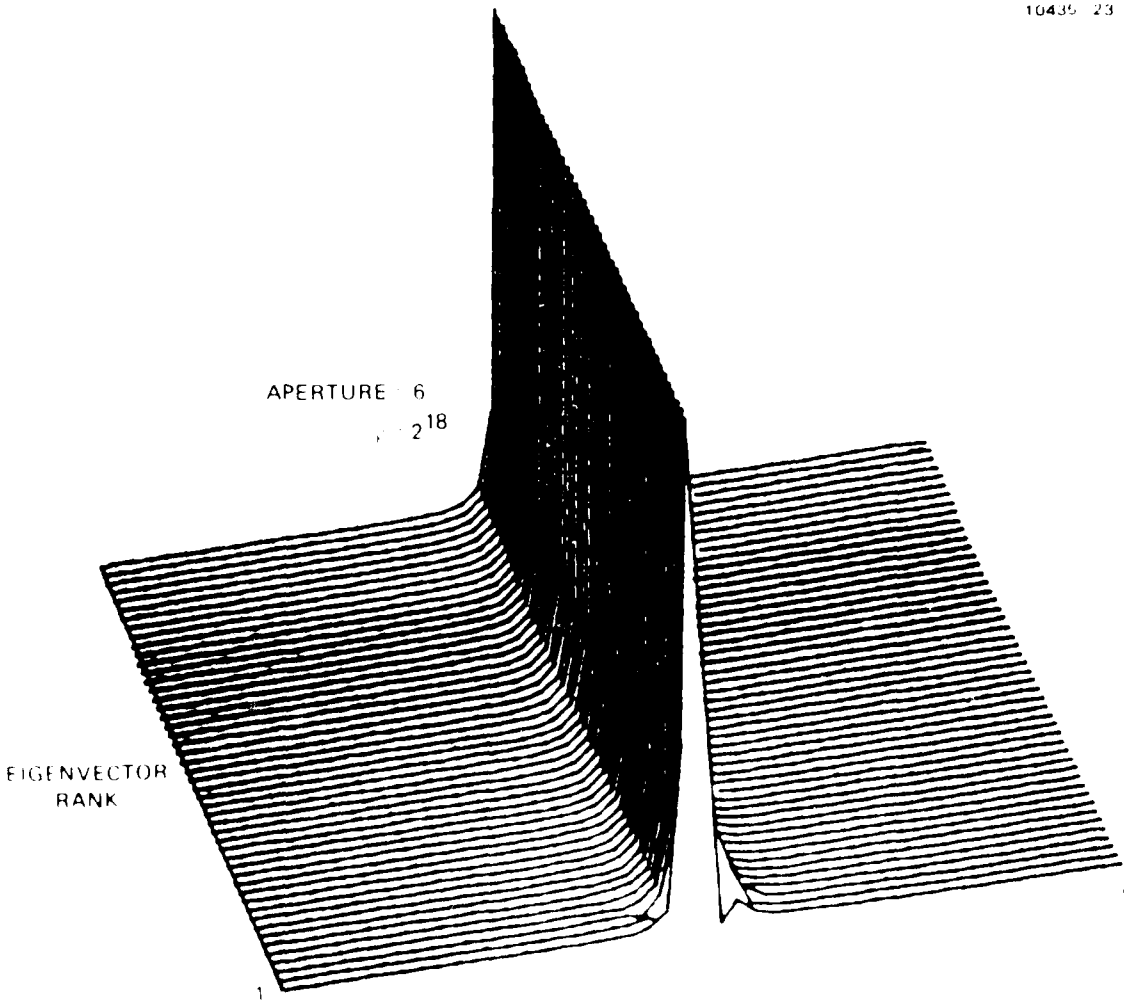
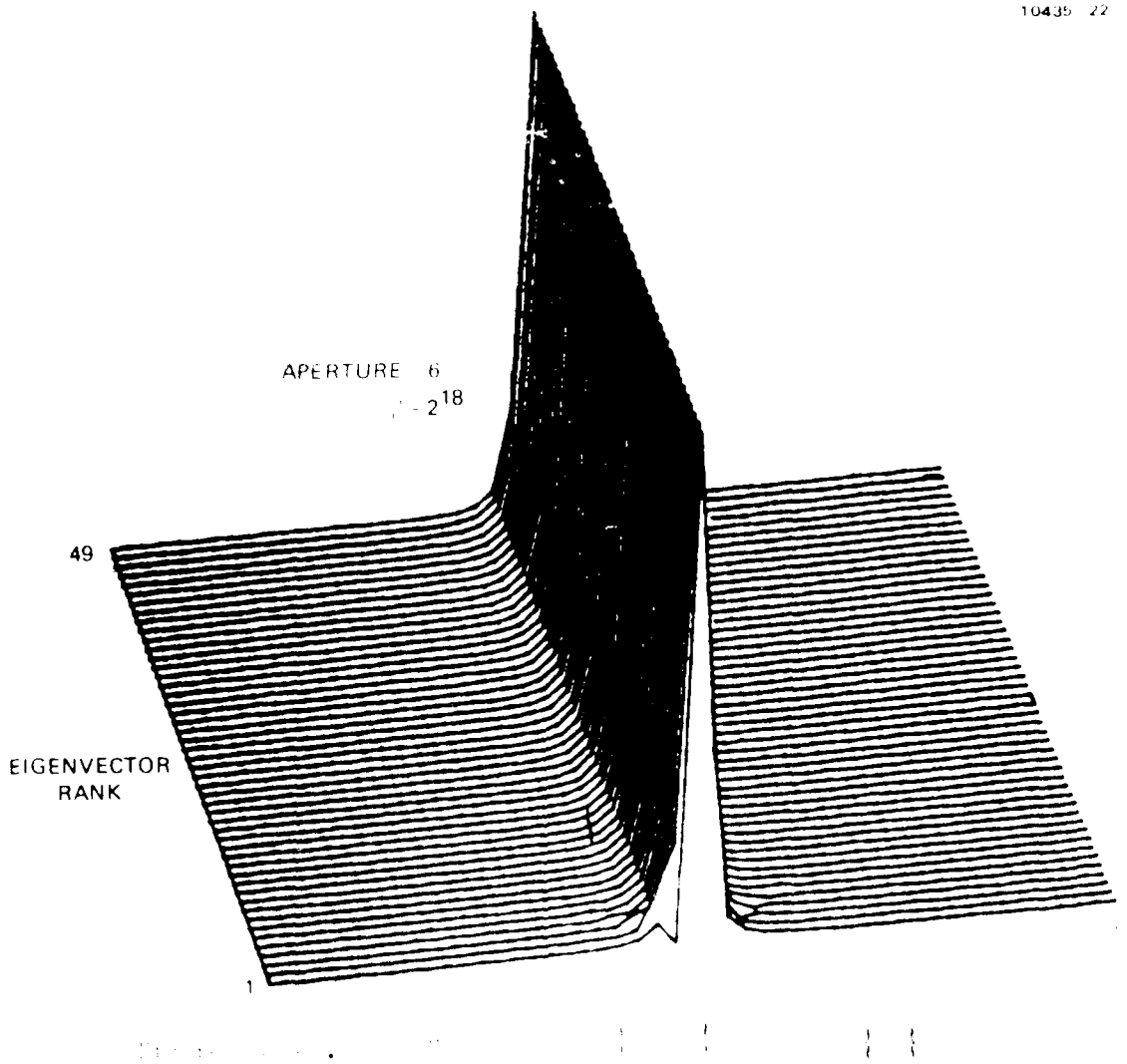


Figure 23(c). Eigenvectors rescaled and projected to p-space as $\sqrt{2}\{a_i\}$ for aperture 6, $\beta = 2^{18}$.


APERTURE 6
2^{18}
EIGENVECTOR
RANK
1


APERTURE 6
2^{18}
49
EIGENVECTOR
RANK
1

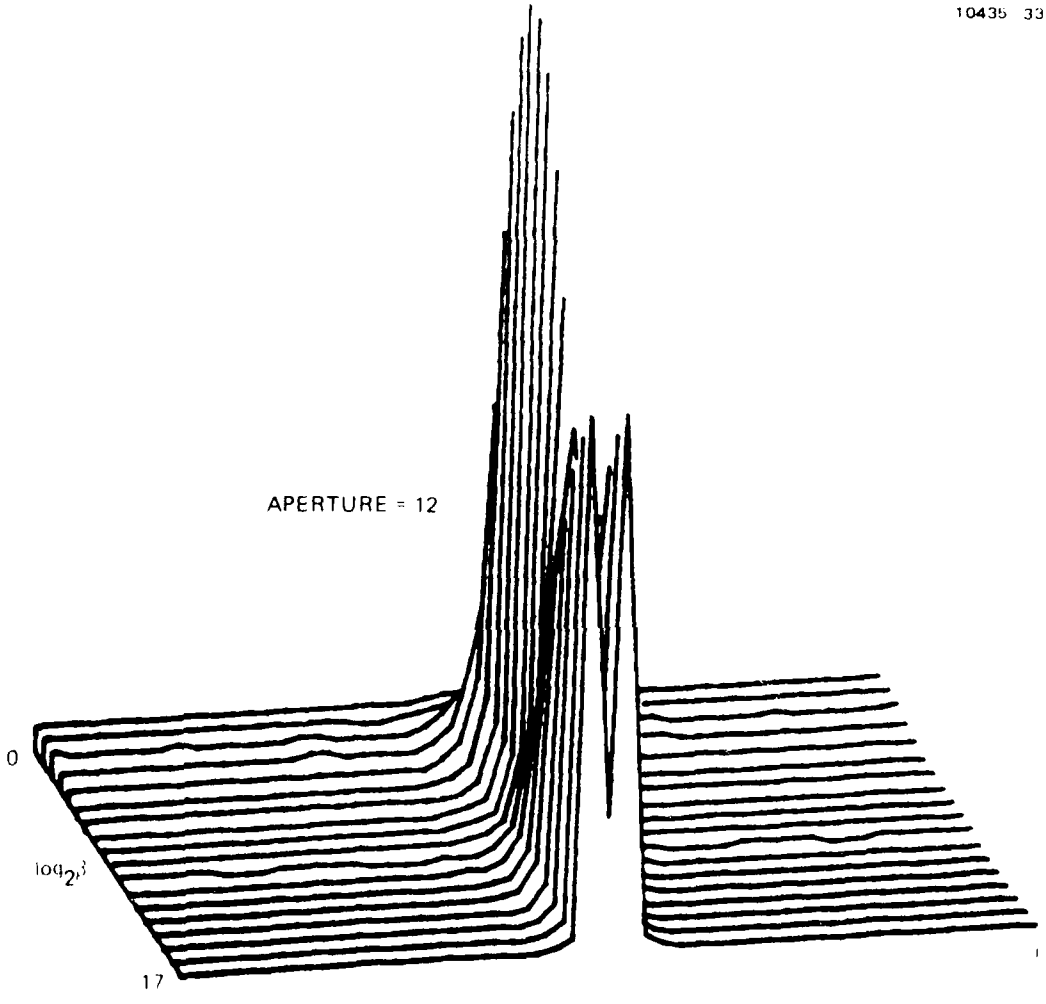


Figure 24. The ME estimates $\{p_i^{(o)}\}_{i=1}^{49}$ for aperture size 12 and for $\log_2 \beta$ = [illegible]

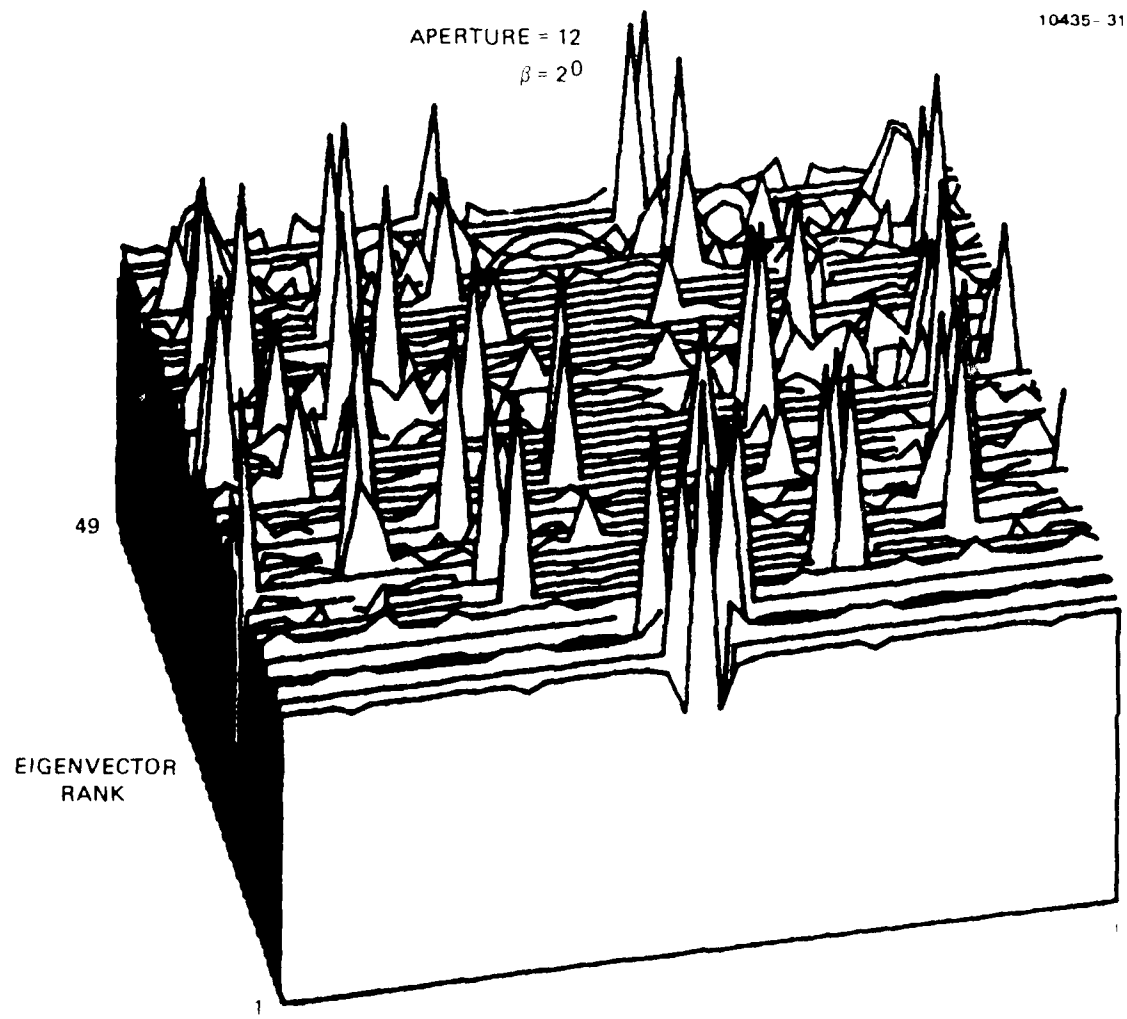


Figure 25(a). Ortho-normal eigenvectors projected on p-space for aperture size 12 and $\beta = 2^0$. Positive peaks.

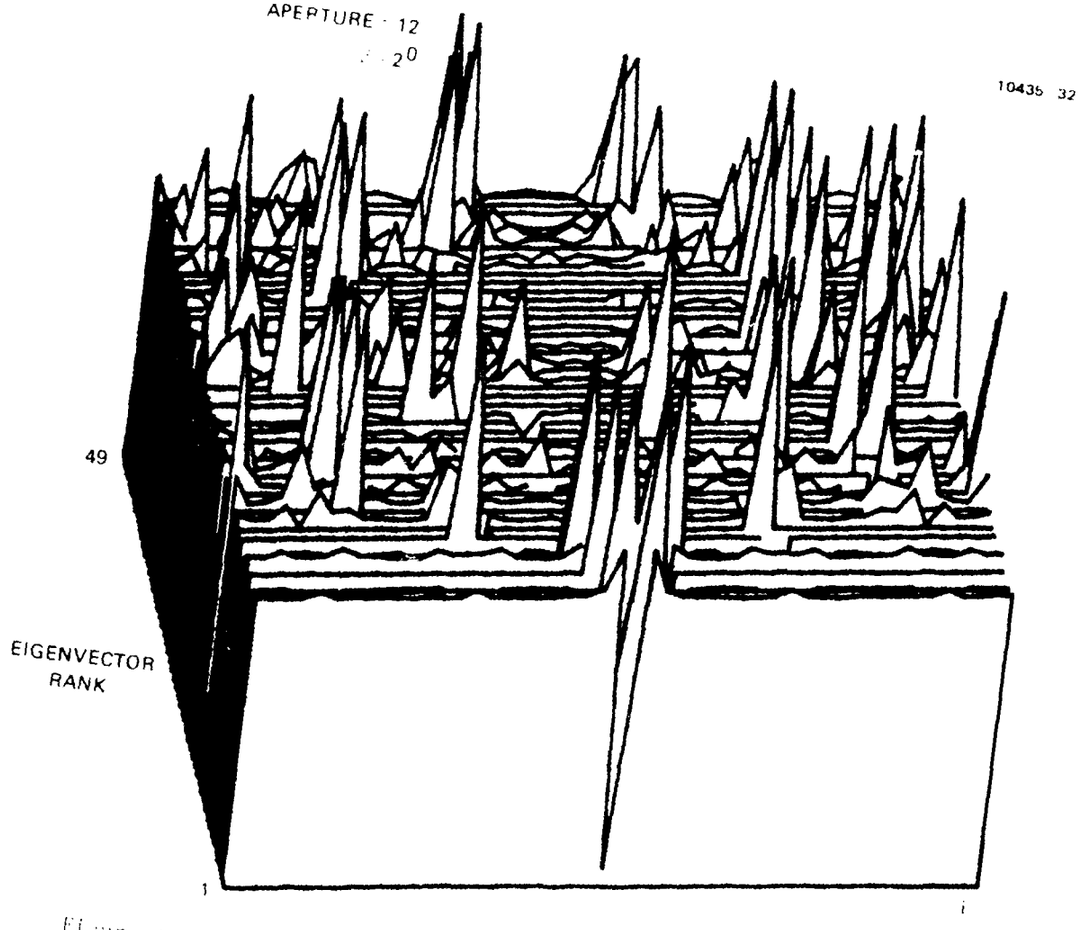


Figure 3.4(b). Ortho-normal eigenvectors projected on p-space for aperture size 12 and $\gamma = 2°$. Negative peaks.

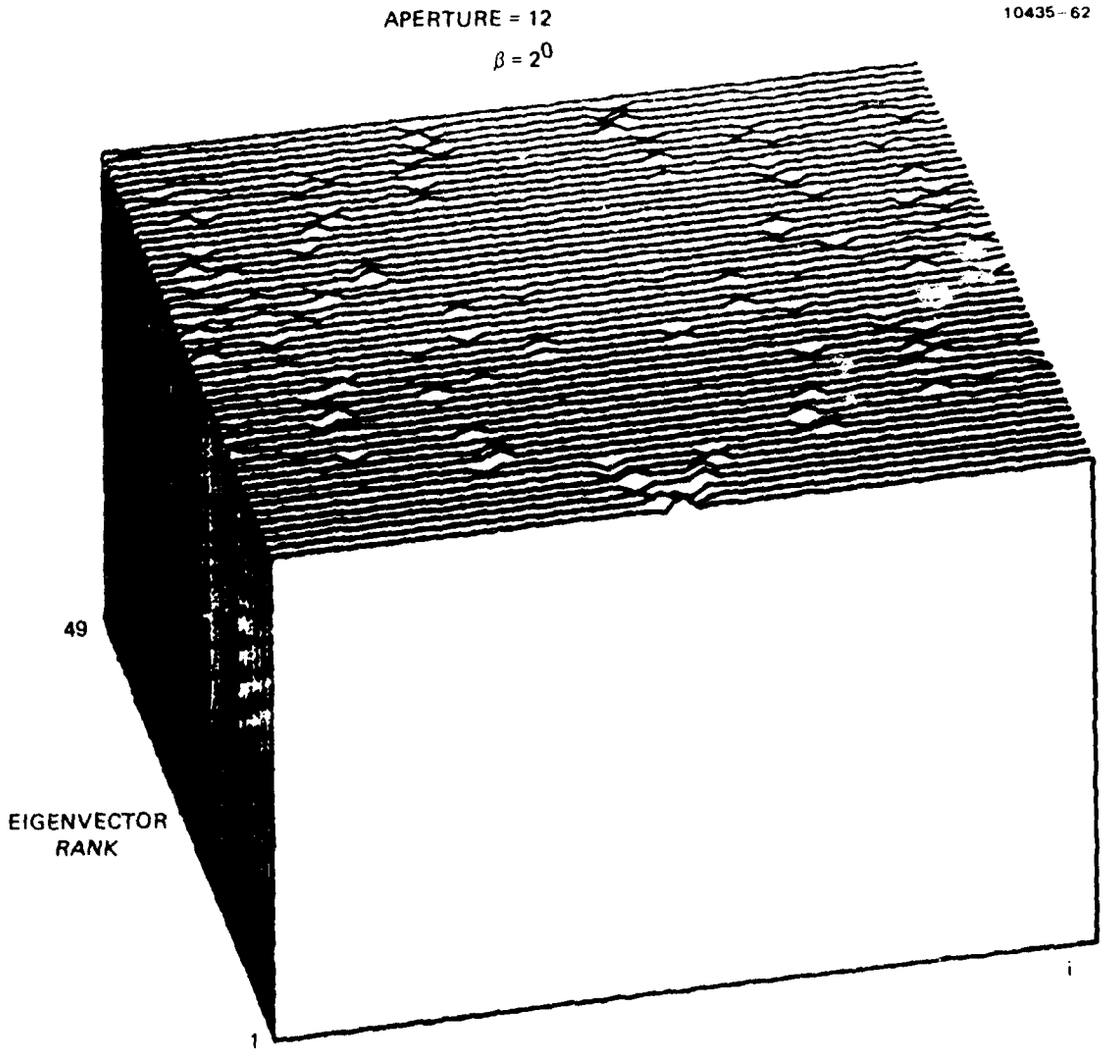


Figure 25(c). Eigenvectors rescaled and projected to p-space as $1.04\{\sigma_i\}$ for aperture 12, $\beta = 2°$.

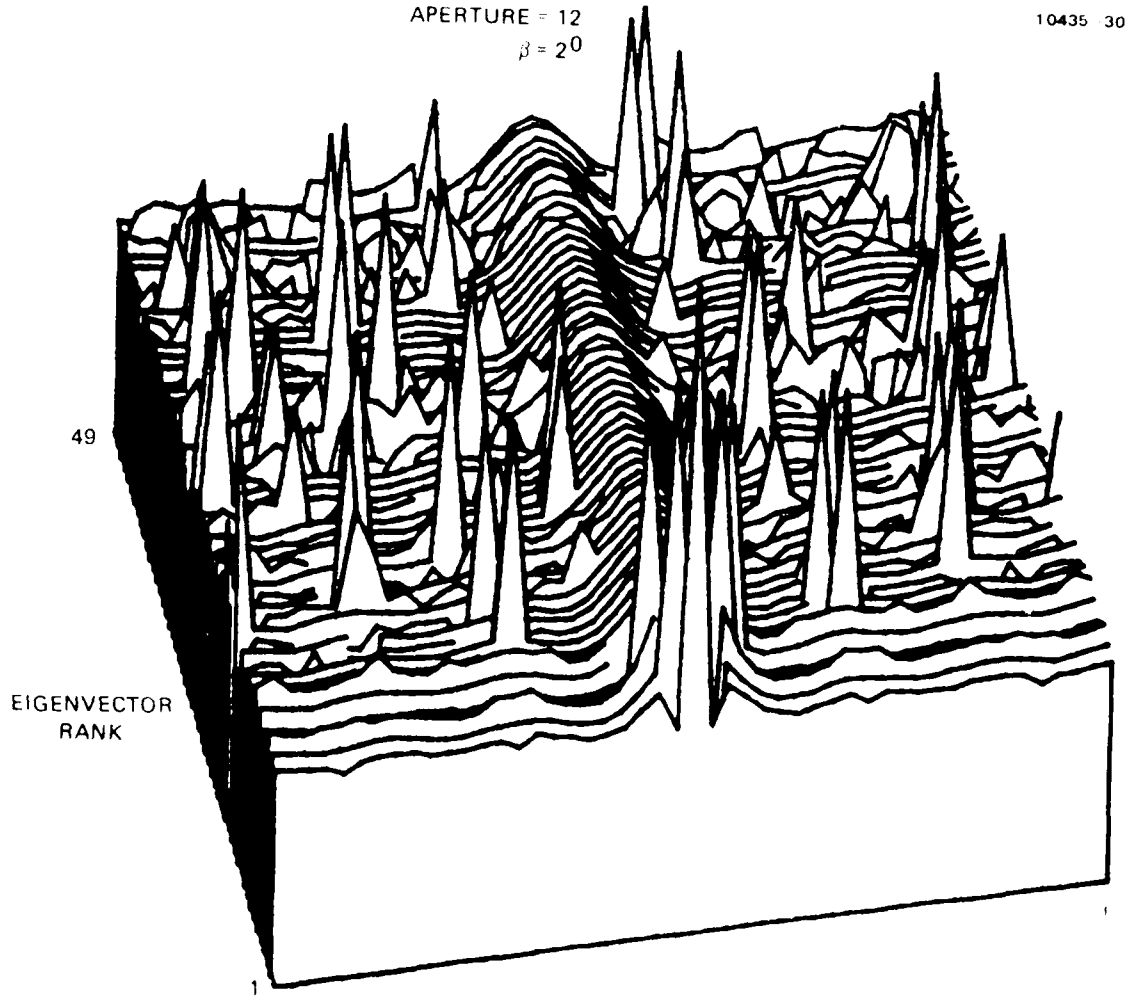


Figure 25(d). The ME estimates $\{p_i^{(o)}\}$ plus 1.04 $\{\tau_i\}$ for aperture 12, $\beta = 2^0$.

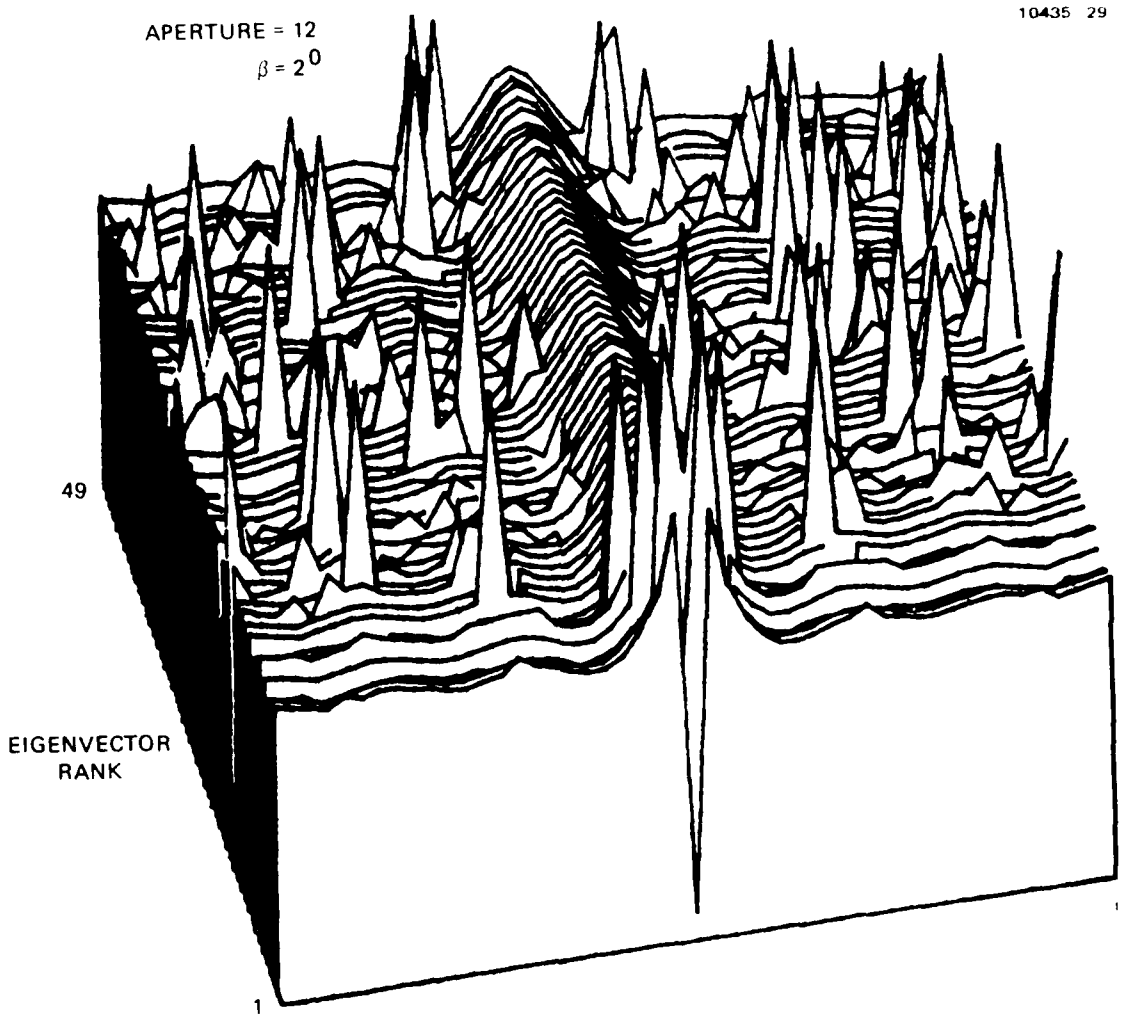


Figure 25(e). The ME estimates $\{p_i^{(o)}\}$ minus 1.04 $\{\ \}$ for aperture 12, $\beta = 2^0$.

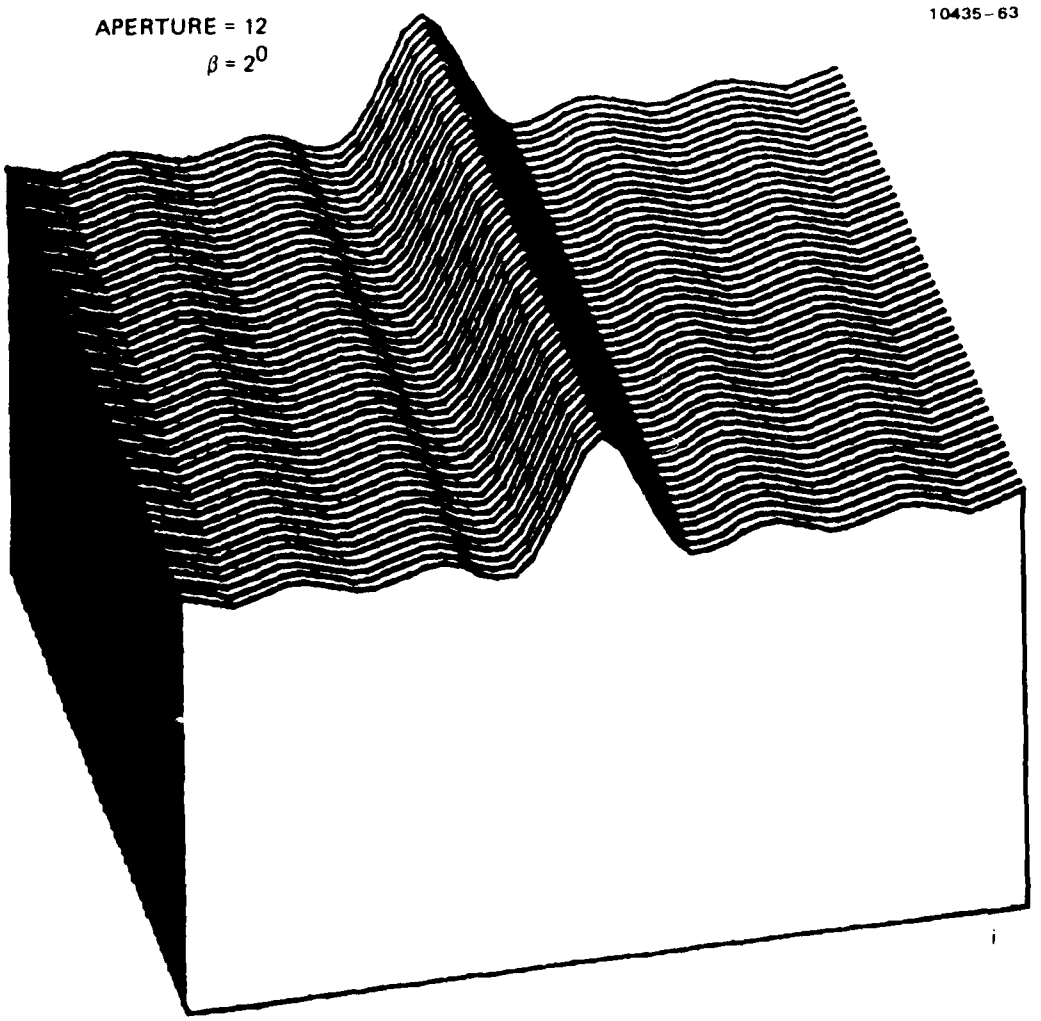


Figure 25(f). The ME estimate $\{p_i^{(o)}\}$ for aperture 12, $\beta = 2°$, replicated 49 times.

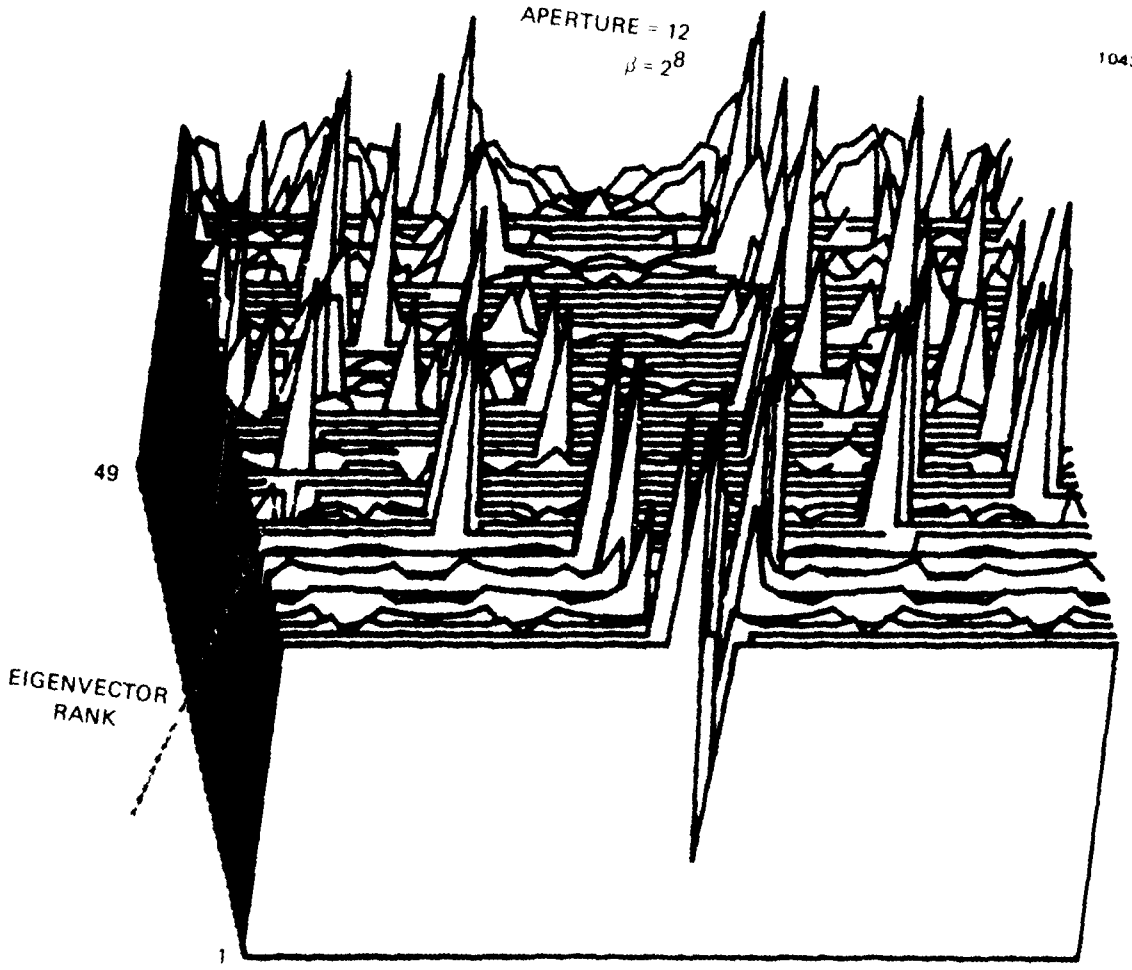


Figure 26(a). Ortho-normal eigenvectors projected on P space for aperture size 12 and $\beta = 2^8$. Positive peaks.

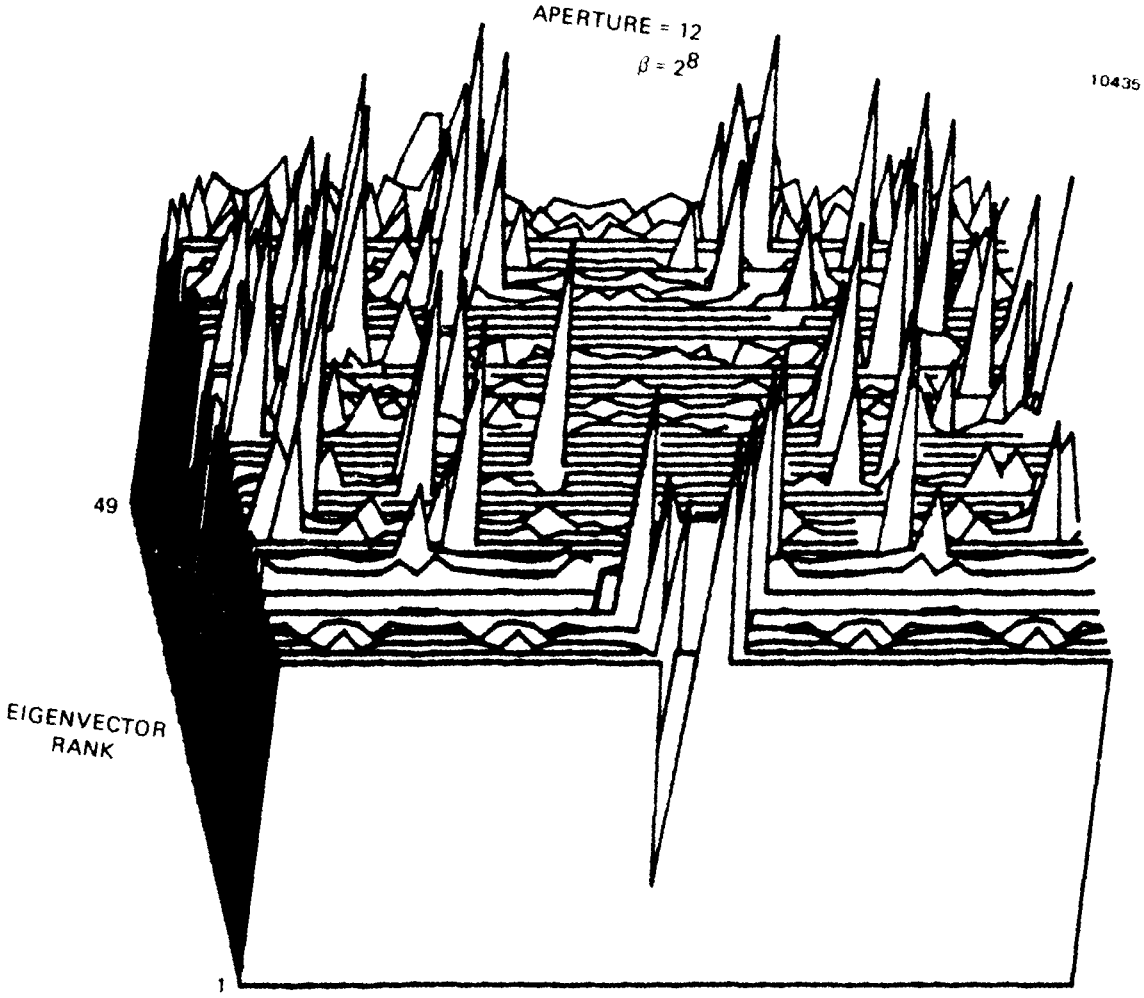


Figure 26(b). Ortho-normal eigenvectors projected on p-space for aperture size 12 and $\beta = 2^8$. Negative peaks.

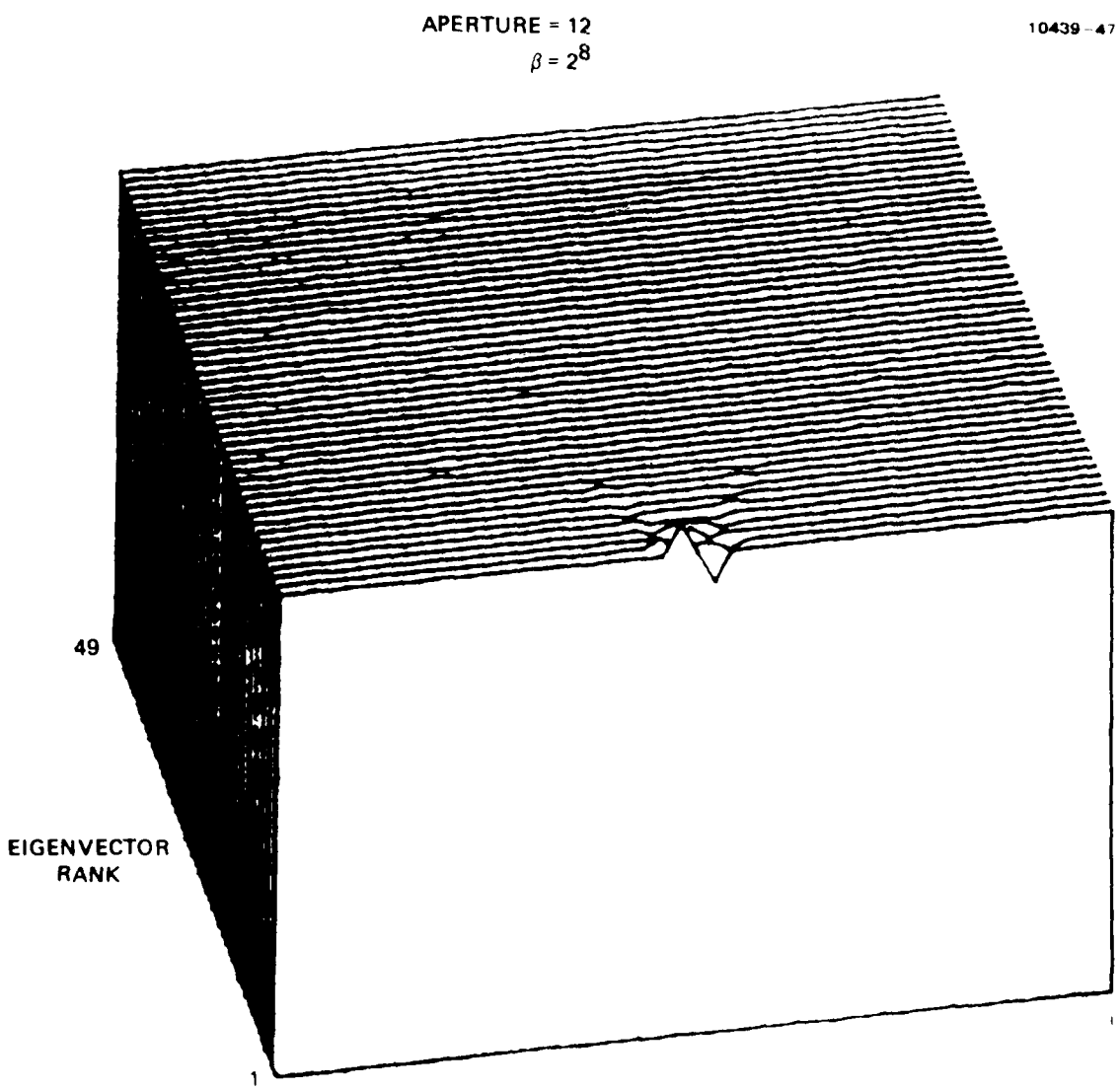
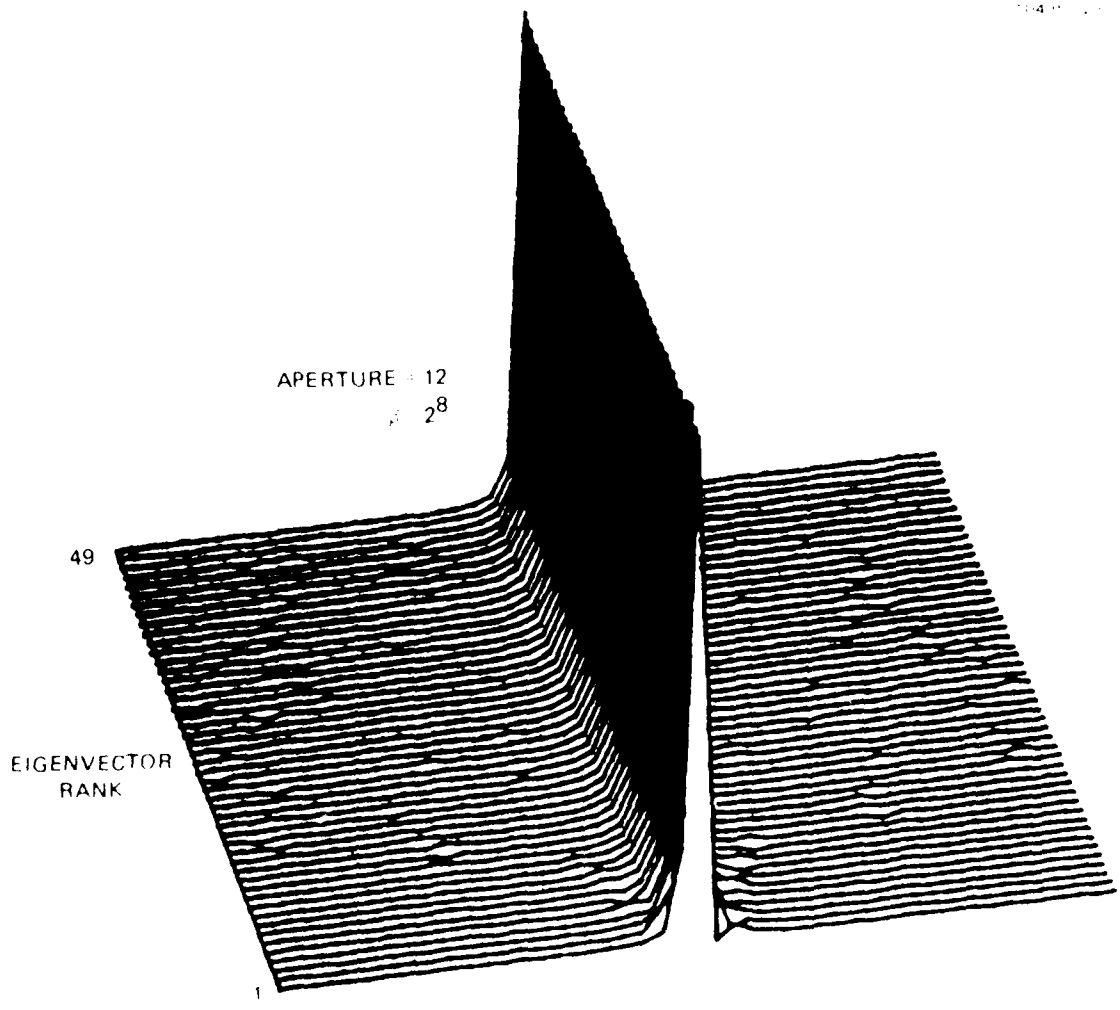
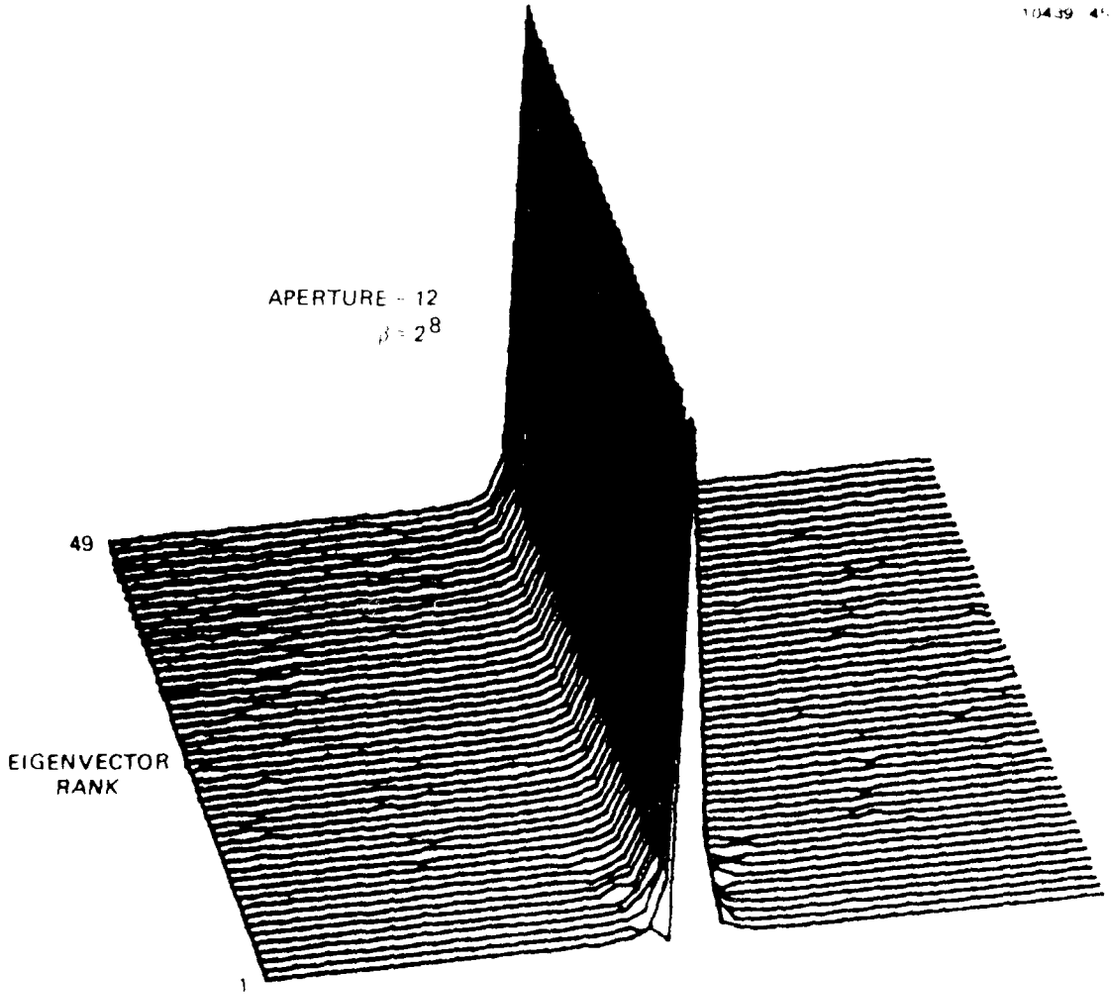


Figure 26(c). Eigenvectors rescaled and projected to p-space [illegible] for aperture 6, $\beta = 2^8$.


APERTURE = 12
2^8
49
EIGENVECTOR
RANK
1


APERTURE = 12
$\beta = 2^8$
49
EIGENVECTOR
RANK
1

one eigenvector has any influence is shown in Figure 27(c) where the base represents minus one. The estimates $\{p_i^{(o)} + \sqrt{2}\}_i$ seen in Figures 27(d) and (e) show the effect of the first eigenvector in putting a confidence band at twice the estimated peak heights.

ME estimates for the case of aperture size 24, for $2^0 \le \gamma \le 2^{19}$ are shown in Figure 28 from the rear, to better show that the ME estimates are beginning to resolve the two peaks when $\gamma \approx 2^6$. It is clear in this reverse plot that for small γ the estimates $\{p_i^{(o)}\}$ tend to a flat distribution of uniform height except for a small single unresolved peak. The confidence estimate is developed in Figures 29(a) through (e) for the case $\gamma = 2^0$. The spread in eigenvalues is slowly increasing as the aperture enlarges. Here, for aperture 24, the ratio of σ is $\lambda_1/\lambda_{49} = 1.47$, still showing that all the σ_i are essentially equal. This can be noted in Figure 29(c), the base of which is at minus one and the mesa top at zero. The upper and lower confidence bands are shown in Figures 29(d) and (e), respectively, with the bases set to zero level. The estimate itself $\{p_i^{(o)}\}$ is shown separately in Figure 29(f) with the base at zero level.

For the case $\gamma = 2^6$, aperture 24, just before obvious superresolution, the confidence results are developed in Figures 30(a) through (e). The estimate $\{p_i^{(o)}\}$ is shown separately in Figure 30(f) with zero at the base. Here the ratio of the first two σ_i is $\lambda_1/\lambda_2 = 9.4$, and the first two eigenvectors explain most of the variance as seen in Figure 30(c), where the base is equal to minus one. The upper and lower bounds of the confidence bands are seen in Figures 30(d) and (e).

For $\gamma = 2^{19}$ aperture 24, see Figure 31. As might be expected, almost all the variance is in one $\sigma_i^2 = \sigma_1^2$. From Figure 31(c) we note that the confidence band will be large only at the two resolved peaks as seen in Figures 31(d) and (e).

The final example for an almost full aperture of 48 is interesting in that it asks no superresolution of the ME method and reveals the behavior in an unclouded way. The ME estimates $\{p_i^{(o)}\}$ themselves are

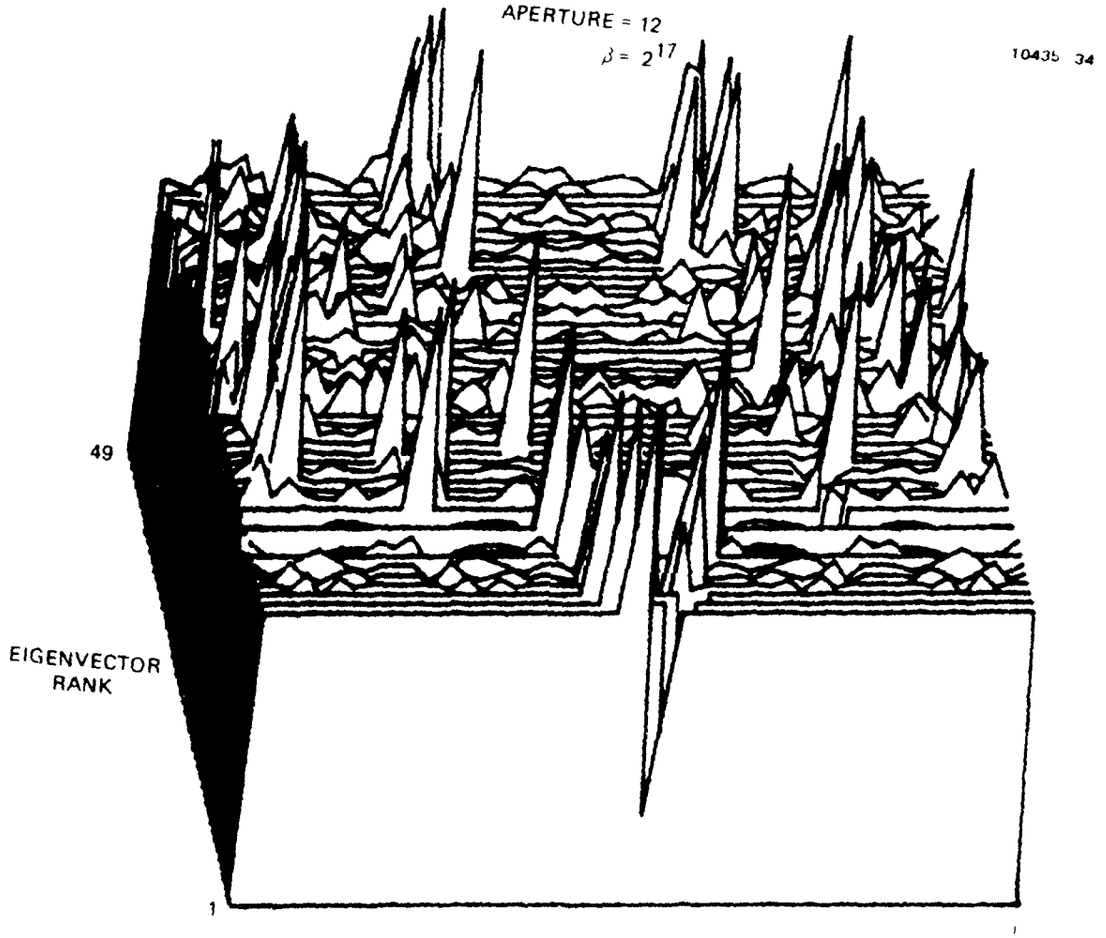


Figure 27(a). Ortho-normal eigenvectors projected on p-space for aperture size 12 and $\beta = 2^{17}$. Positive peaks.

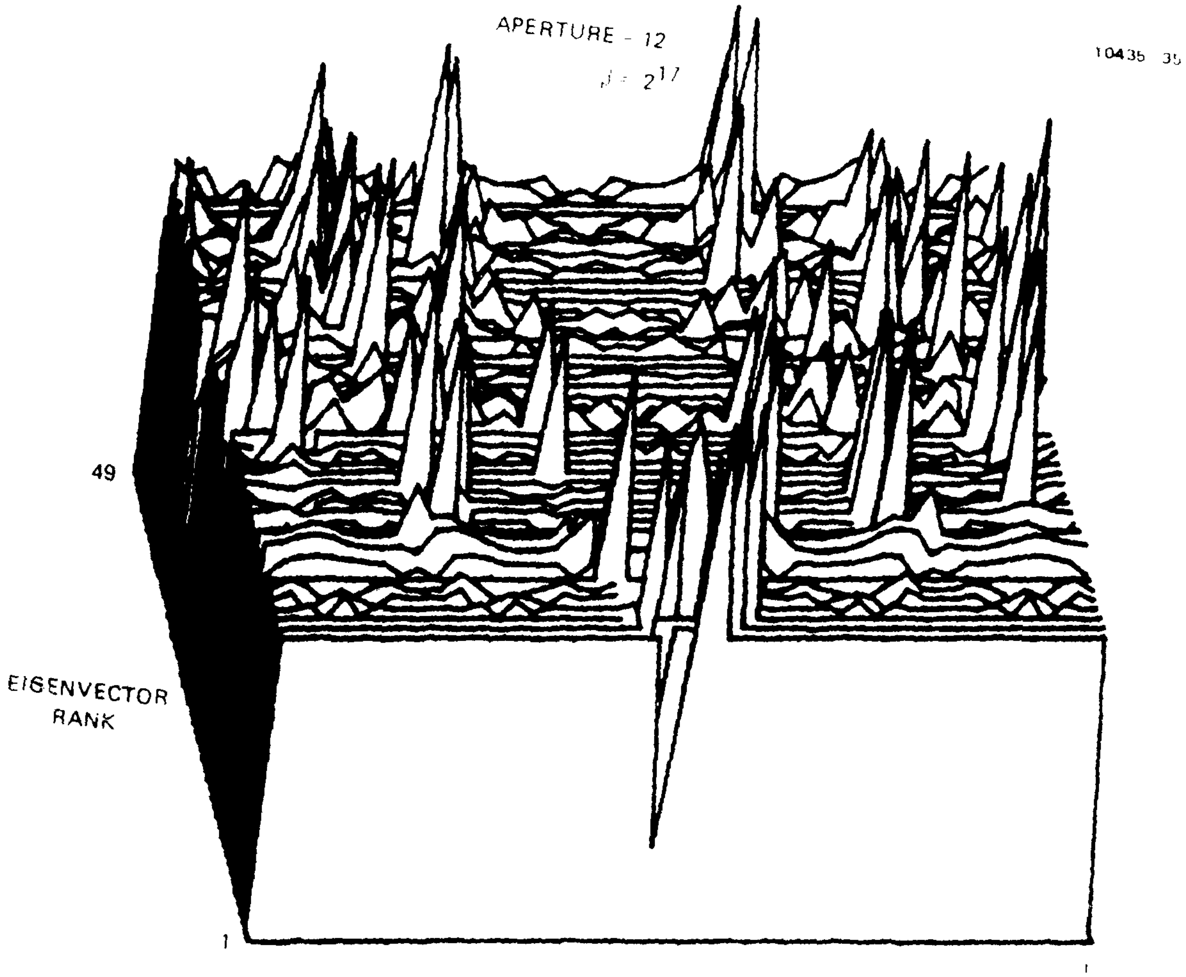


Figure 27(b). Ortho-normal eigenset re-projected on p-space for aperture size 12 and $\beta = 2^{17}$. Negative peaks.

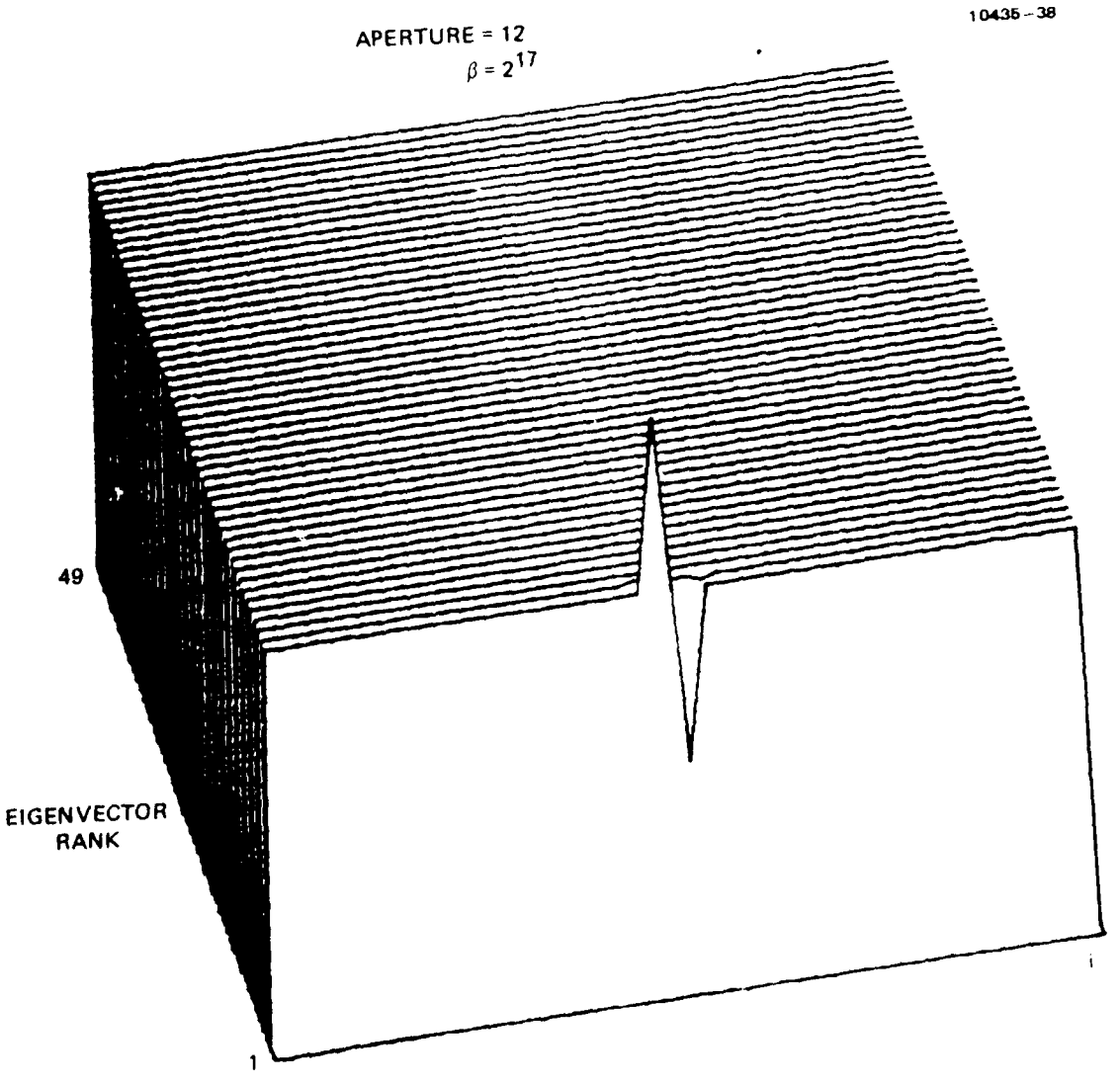
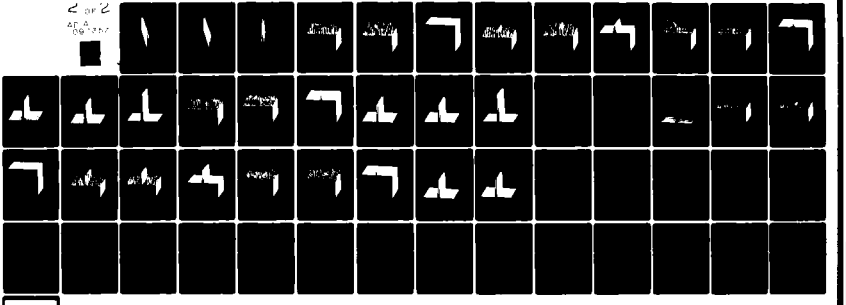


Figure 27(c). Eigenvectors rescaled and projected to p-space as $\sqrt{2}\{\sigma_i\}$ for aperture 12, $\beta = 2^{17}$.

AD-A097 357 HUGHES RESEARCH LABS MALIBU CA F/G 5/8
MAXIMUM ENTROPY IMAGE ESTIMATION ANALYSIS OF IMAGE FORMATION WI--ETC(U)
FEB 81 B H SOFFER, R KIKUCHI F49620-77-C-0052
UNCLASSIFIED AFOSR-TR-81-0324 NL

2 OF 2
AD A 097357



END
DATE FILMED
5-81
DTIC

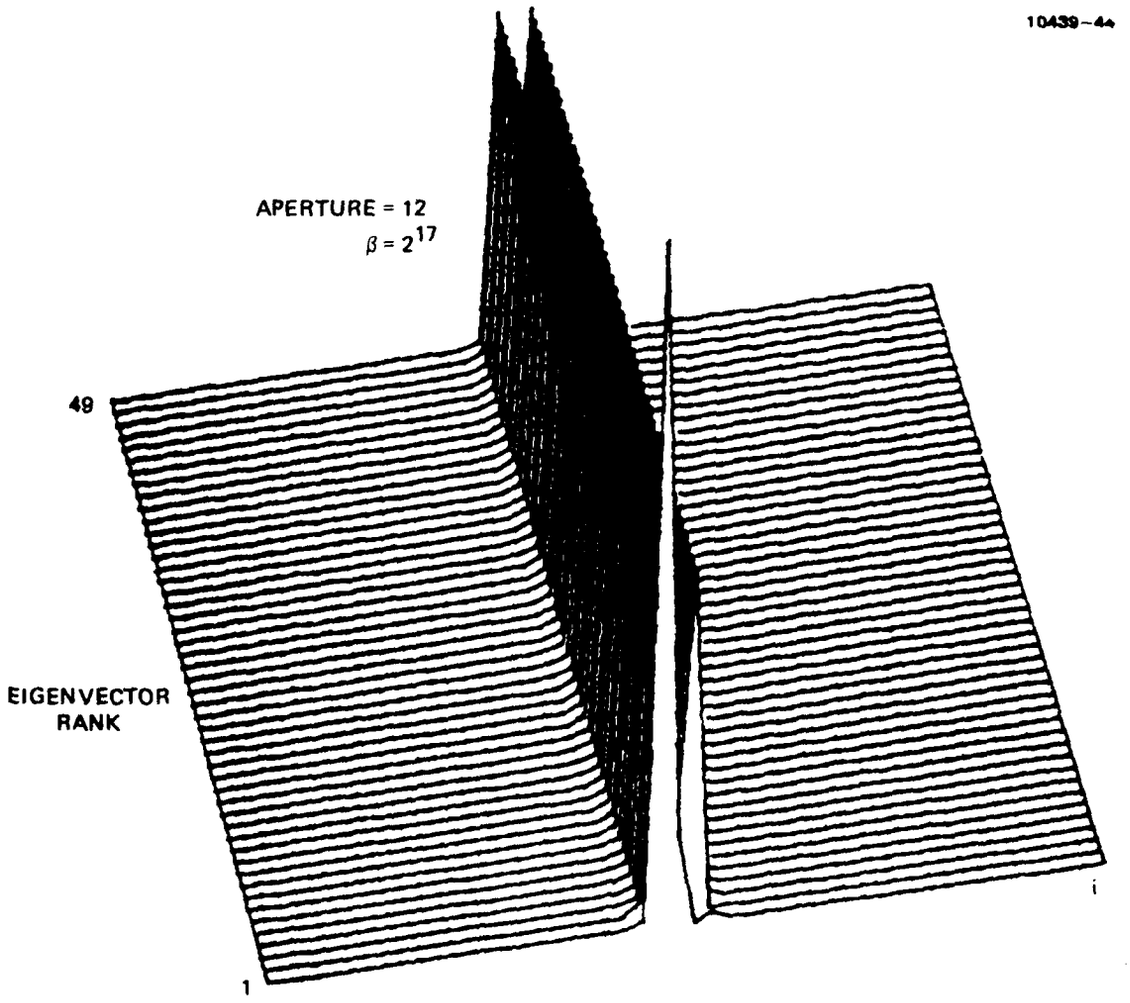


Figure 27(d). The ME estimates $\{p_i^{(o)}\}$ plus $\sqrt{2}$ $\{\sigma_i\}$ for aperture 12, $\beta = 2^{17}$.

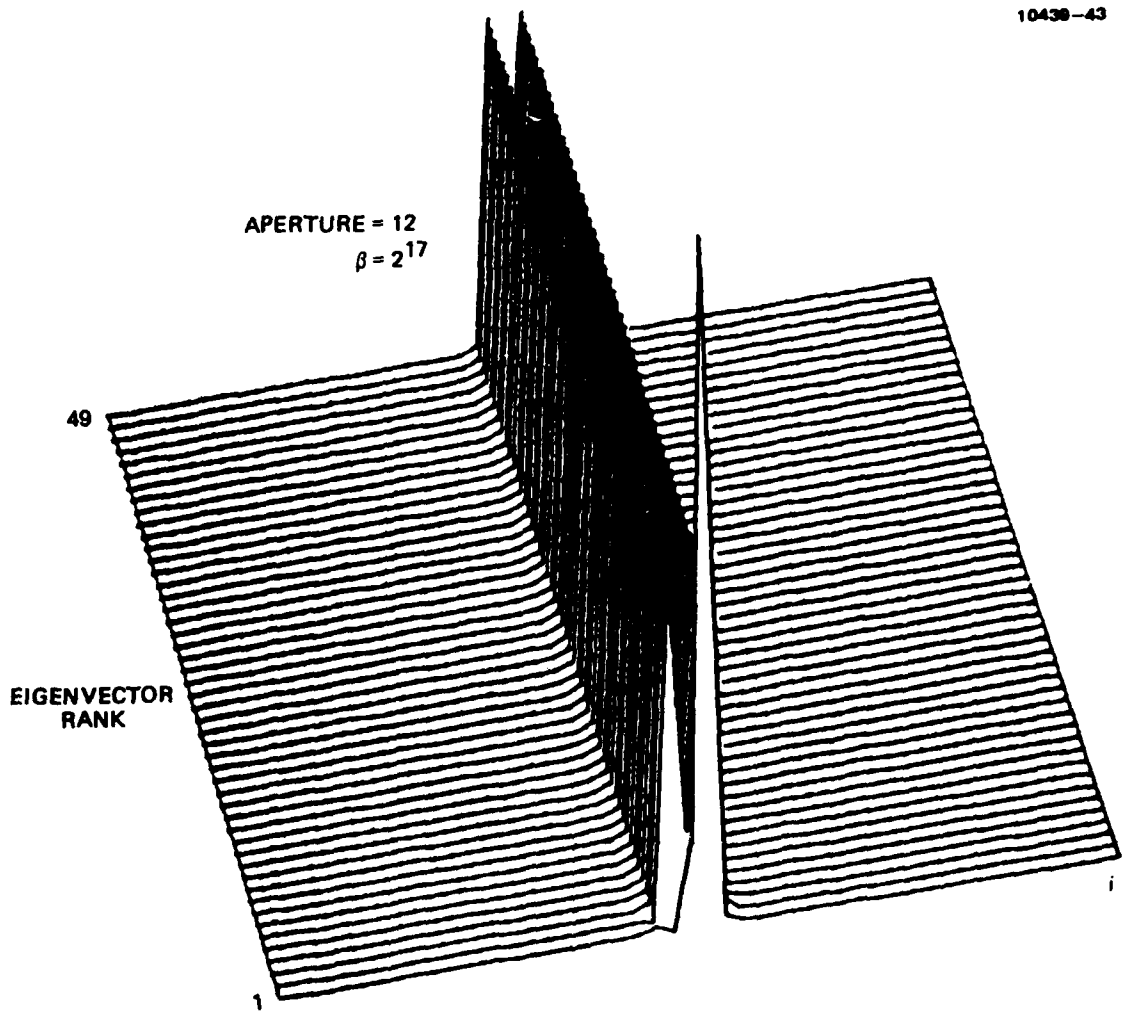


Figure 27(e). The ME estimate $\{p_i^{(o)}\}$ minus $\sqrt{2}\ \{\sigma_i\}$ for aperture 12, $\beta = 2^{17}$.

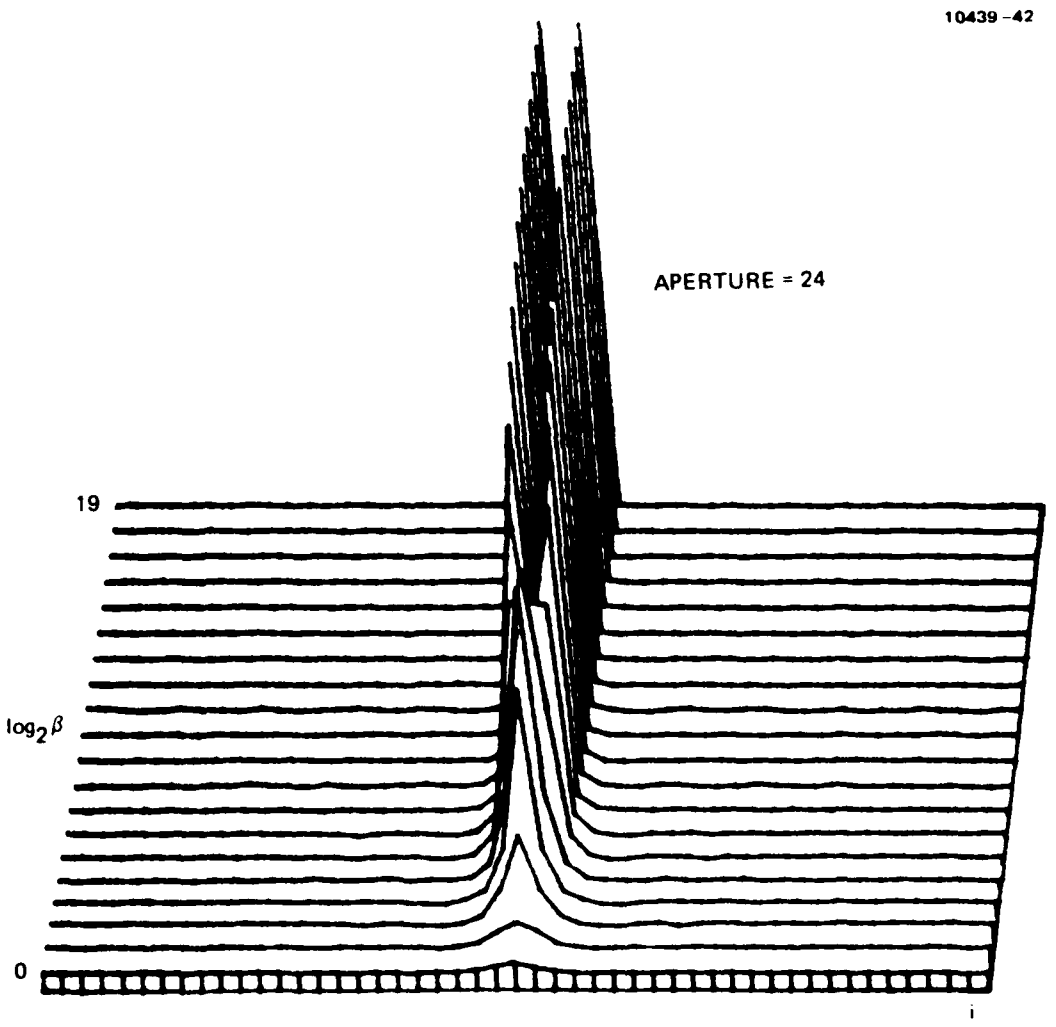


Figure 28. The ME estimates $\{p_i^{(o)}\}$, $1 \le i \le 49$ for aperture size 24 and for $20 < \beta < 219$.

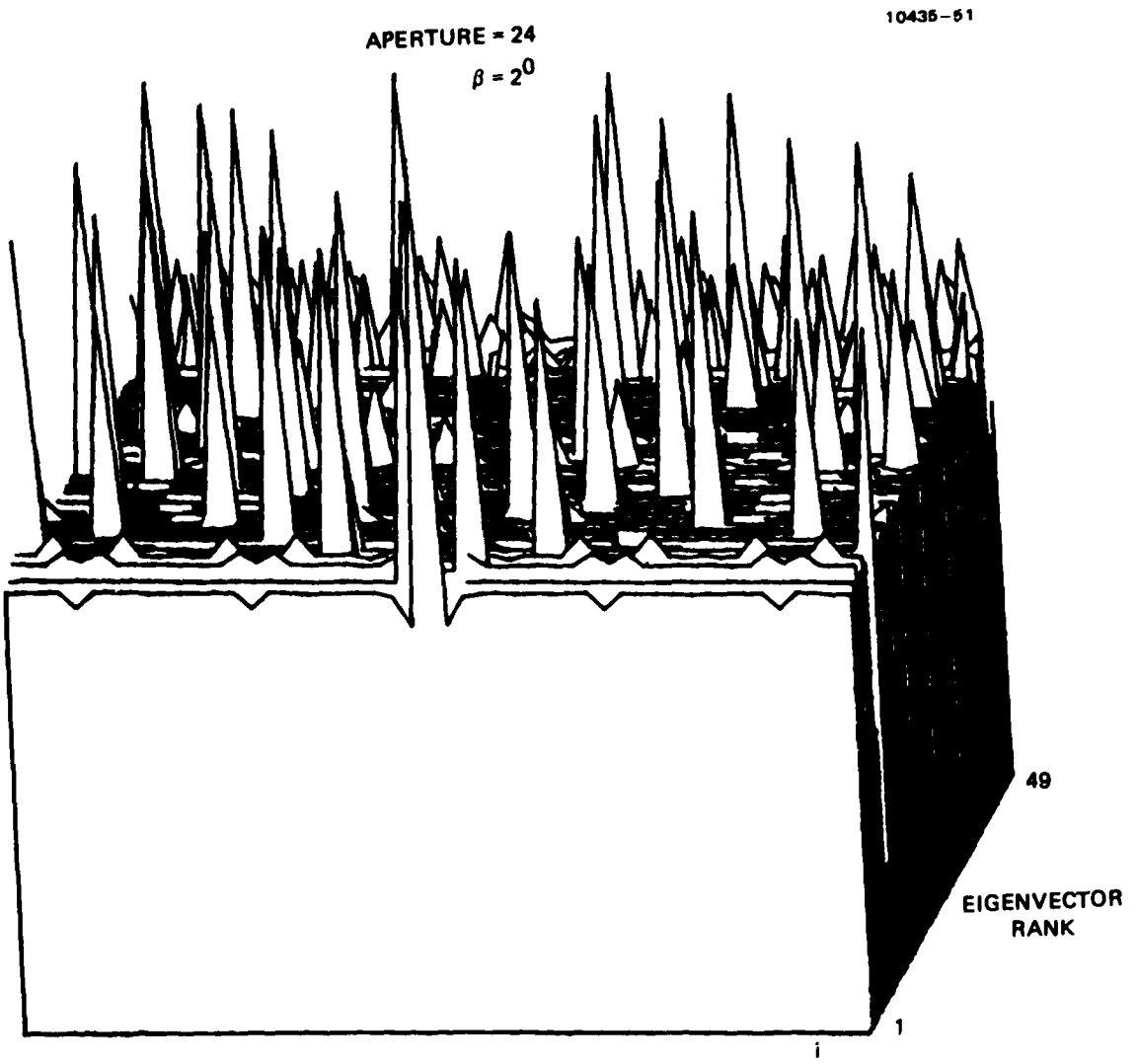


Figure 29(a). Ortho-normal eigenvectors projected on p-space for aperture size 24 and β = 2°. Positive peaks.

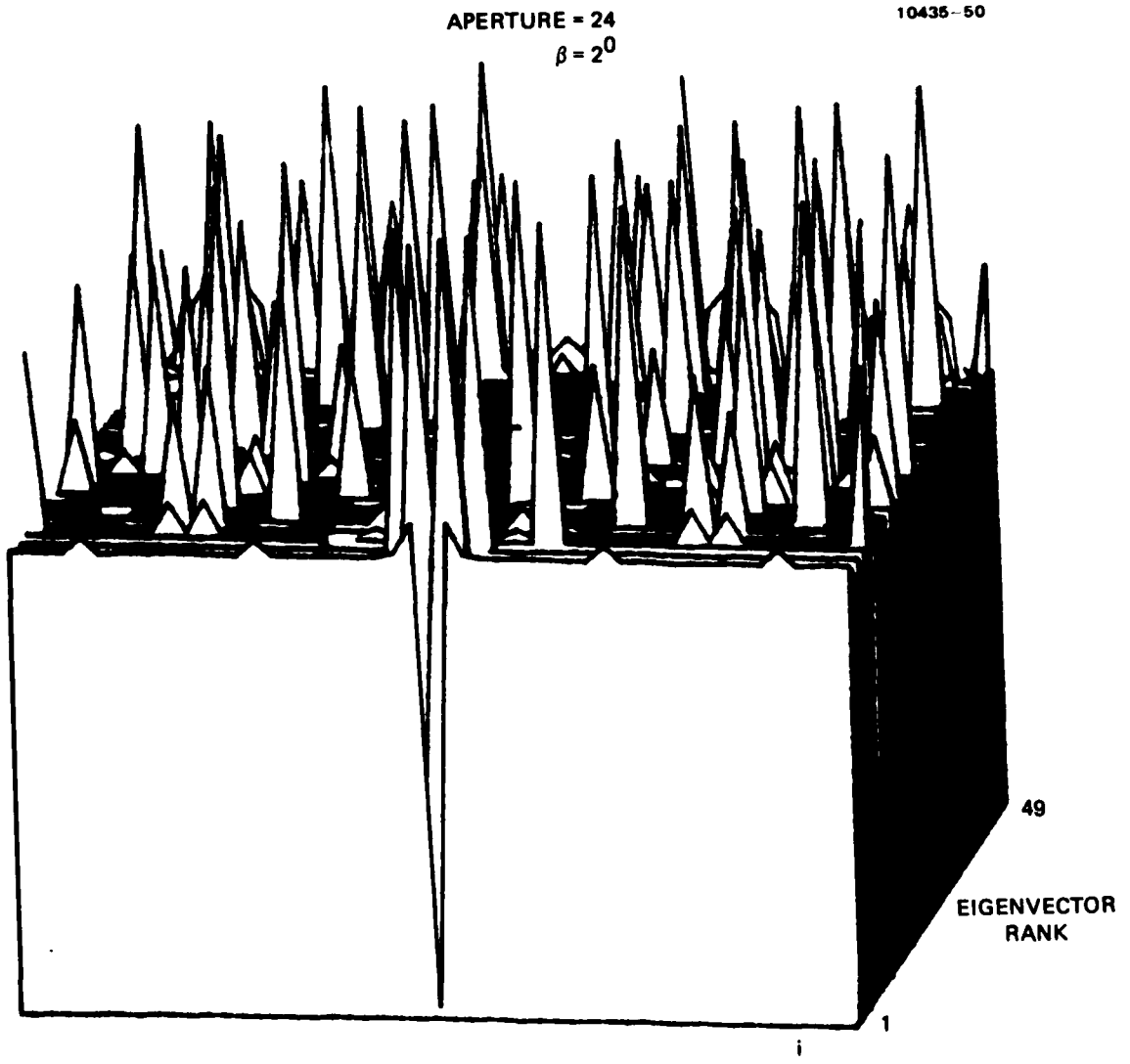


Figure 29(b). Ortho-normal eigenvectors projected for p-space for aperture size 24 and β = 2°. Negative peaks.

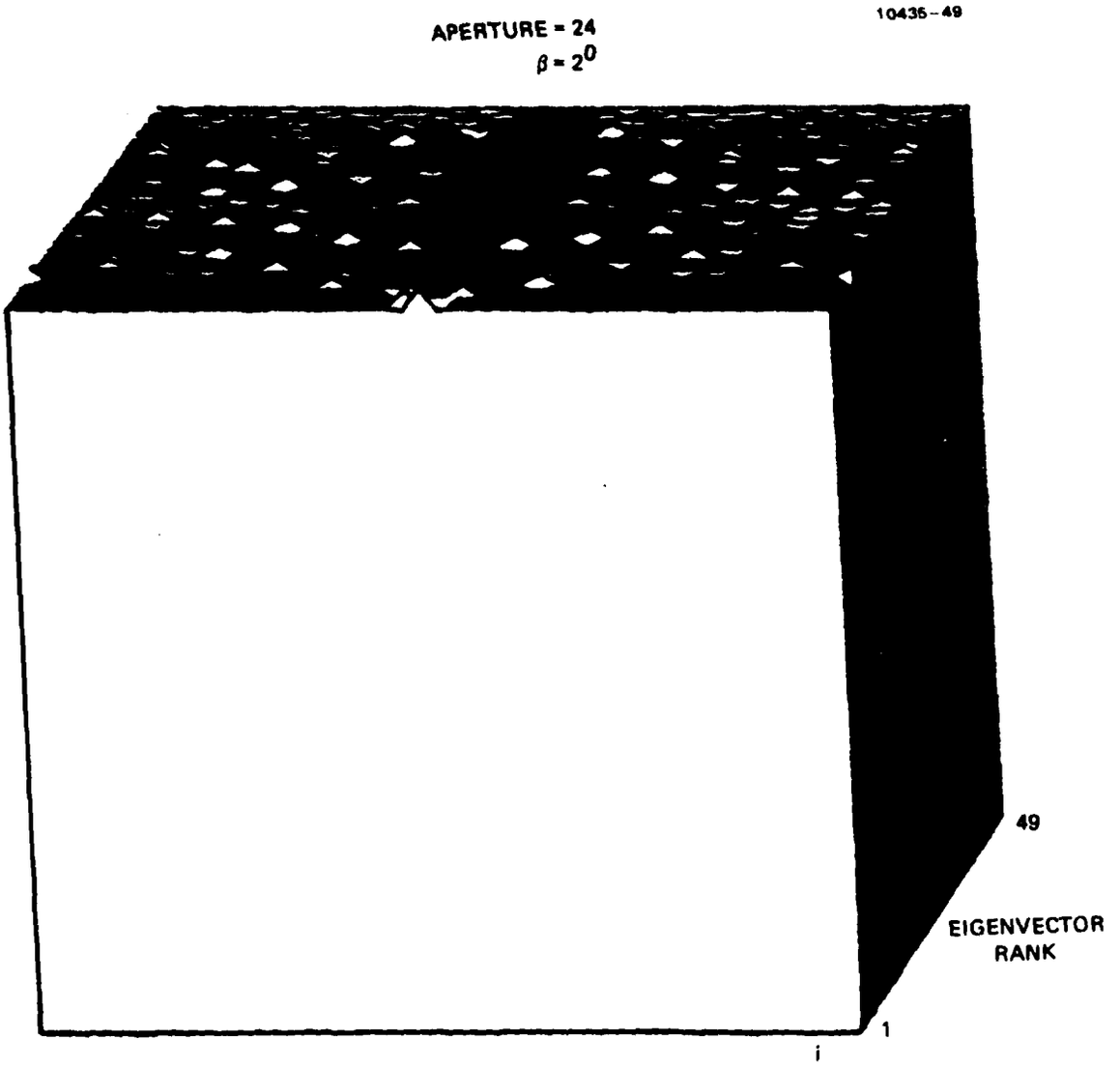


Figure 29(c). Eigenvectors rescaled and projected to p-space as $1.04\ \{\sigma_i\}$ for aperture 24, $\beta = 2°$.

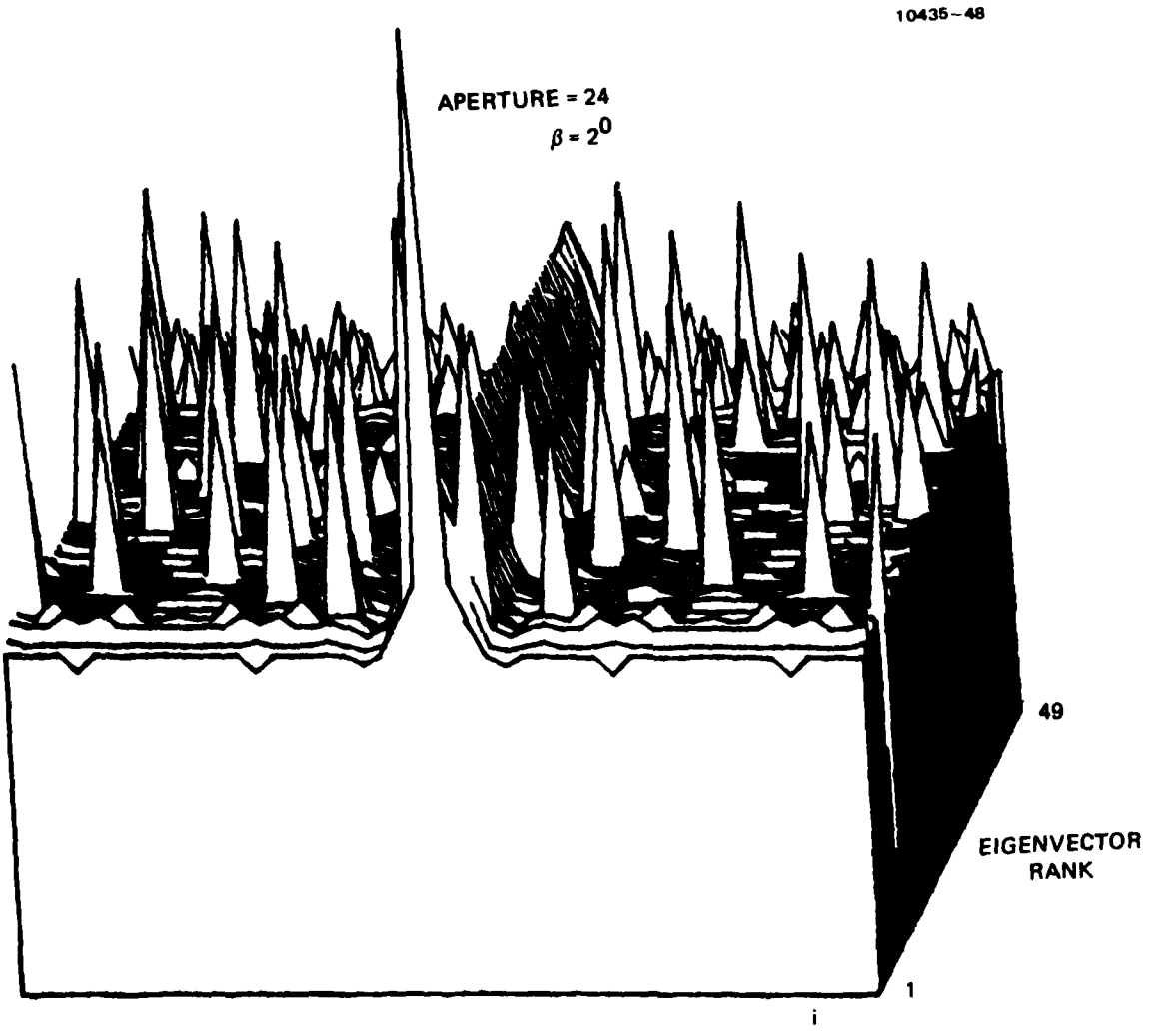


Figure 29(d). The ME estimates $\{p_i(o)\}$ plus 1.04 $\{o_1\}$ for aperture 24, $\beta = 2°$.

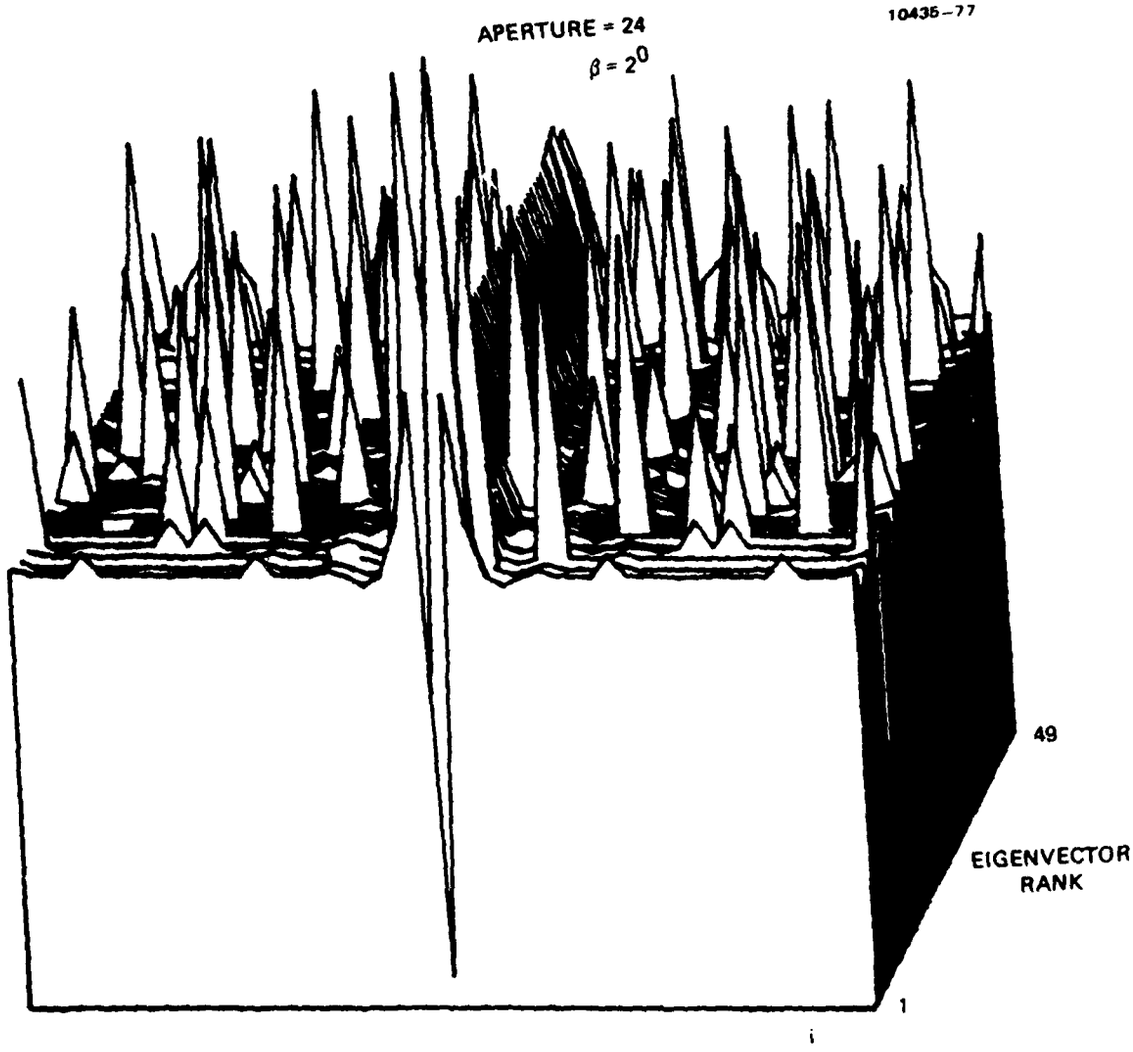


Figure 29(e). The ME estimates $\{p_i^{(o)}\}$ minus 1.04 $\{e_i\}$ for aperture 24, $\beta = 2°$.

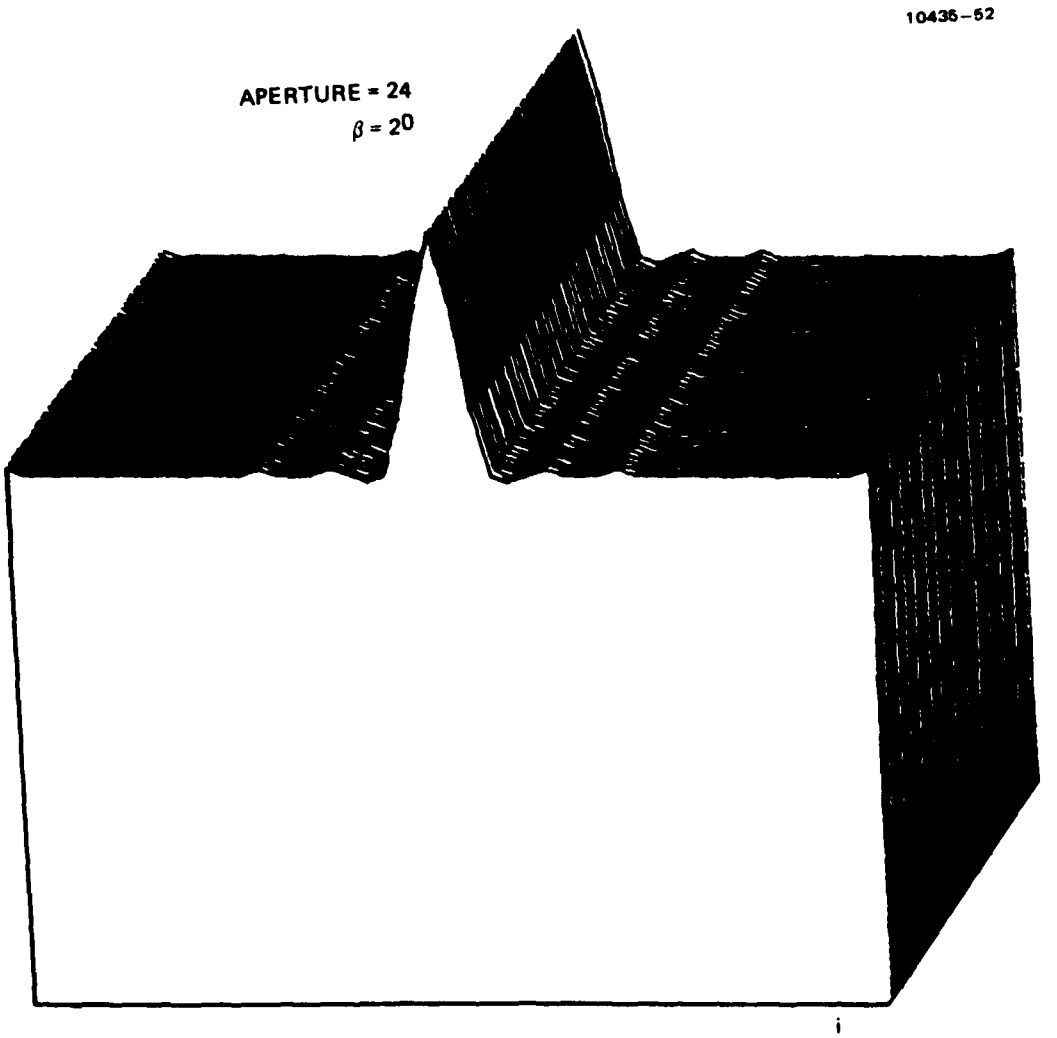


Figure 29(f). The ME estimate $\{p_i^{(o)}\}$ for aperture 24, $\beta = 2^0$, replicated 49 times.

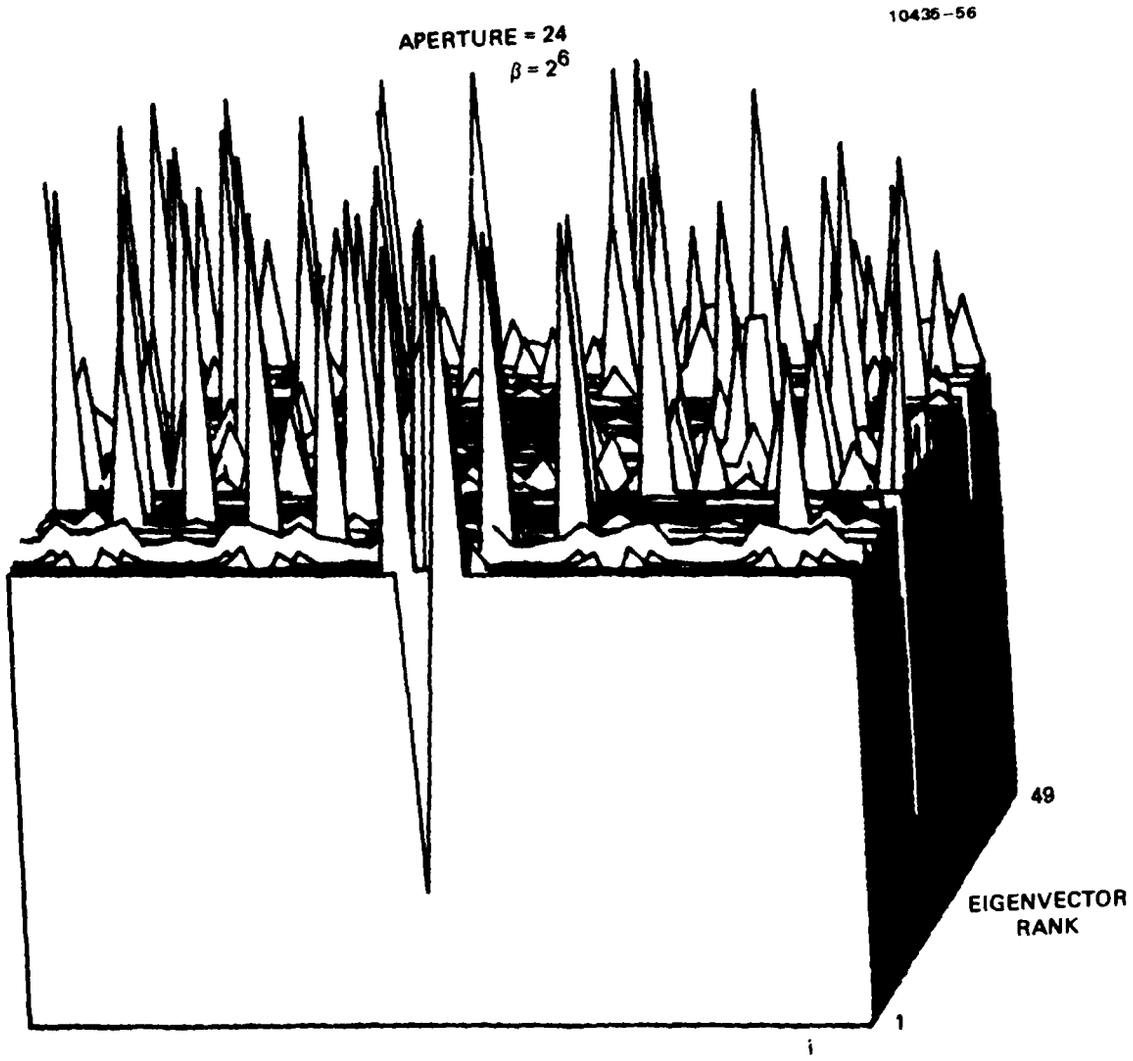


Figure 30(a). Ortho-normal eigenvectors projected on p-space for aperture size 24 and $\beta = 2^6$. Positive peaks.

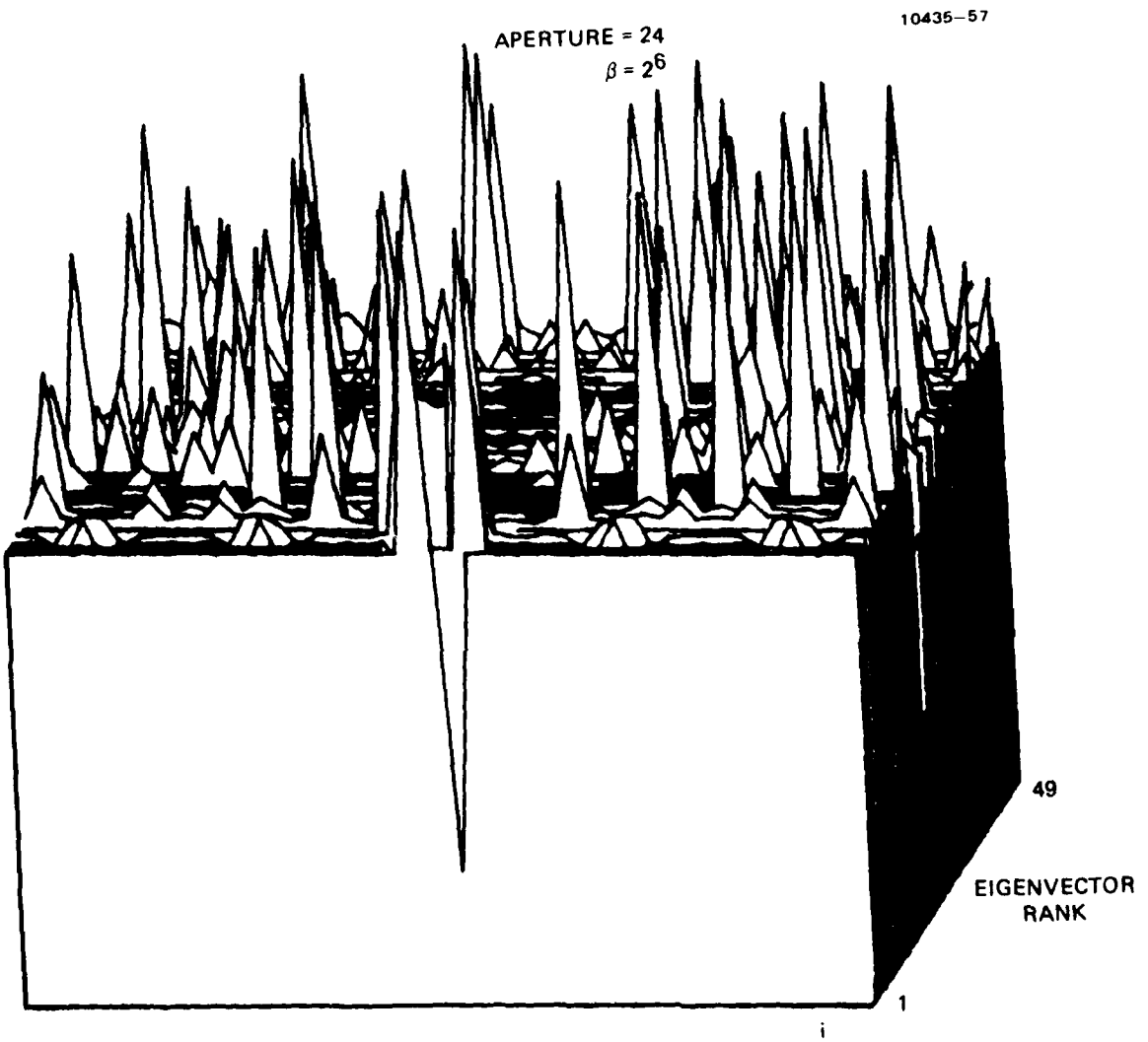


Figure 30(b). Ortho-normal eigenvectors projected on p-space for aperture size 24 and $\beta = 26$. Negative peaks.

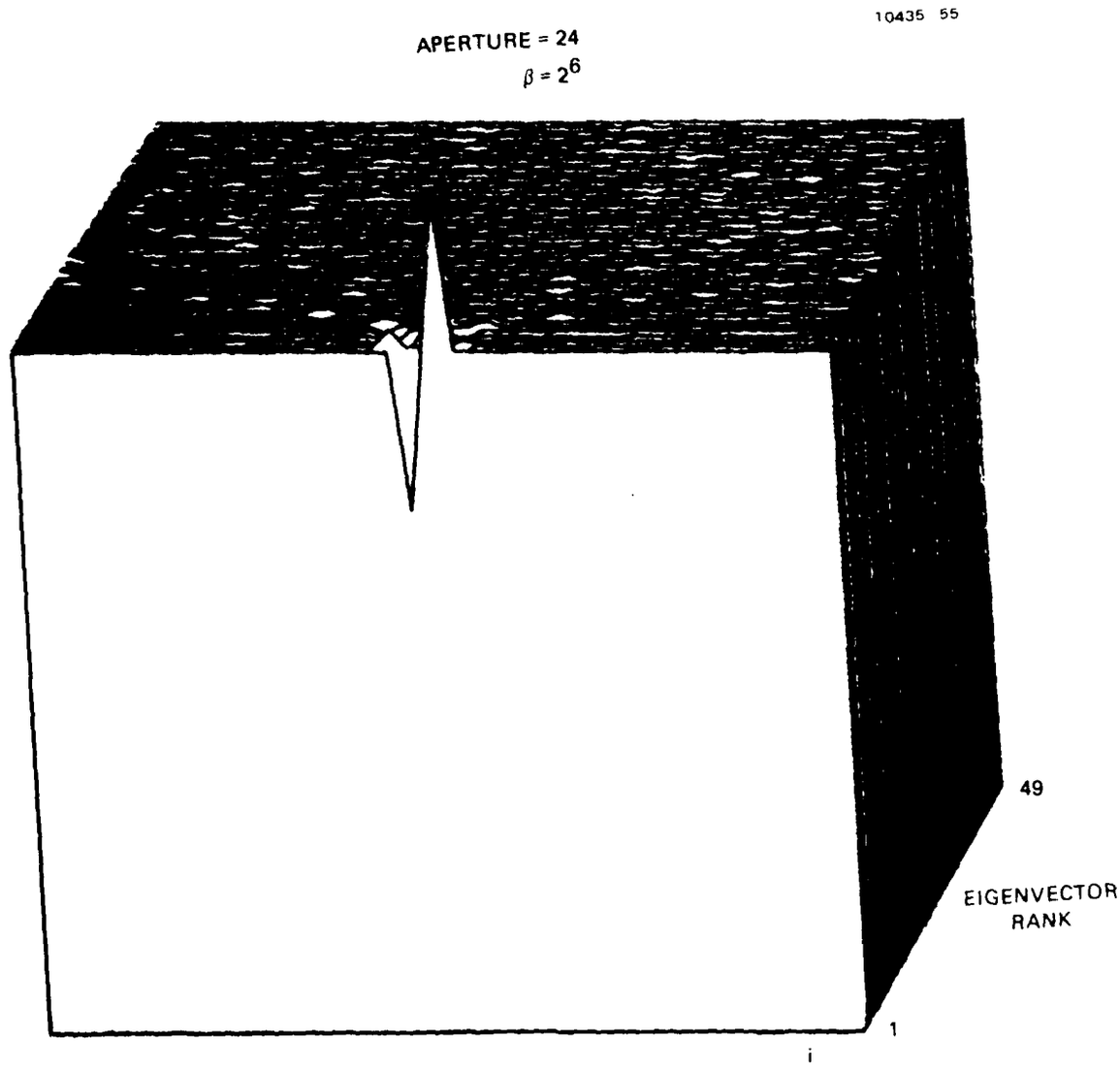


Figure 30(c). Eigenvectors rescaled and projected to p space as $\sqrt{2}(\sigma_i)$ for aperture 24, $\beta = 2^6$.

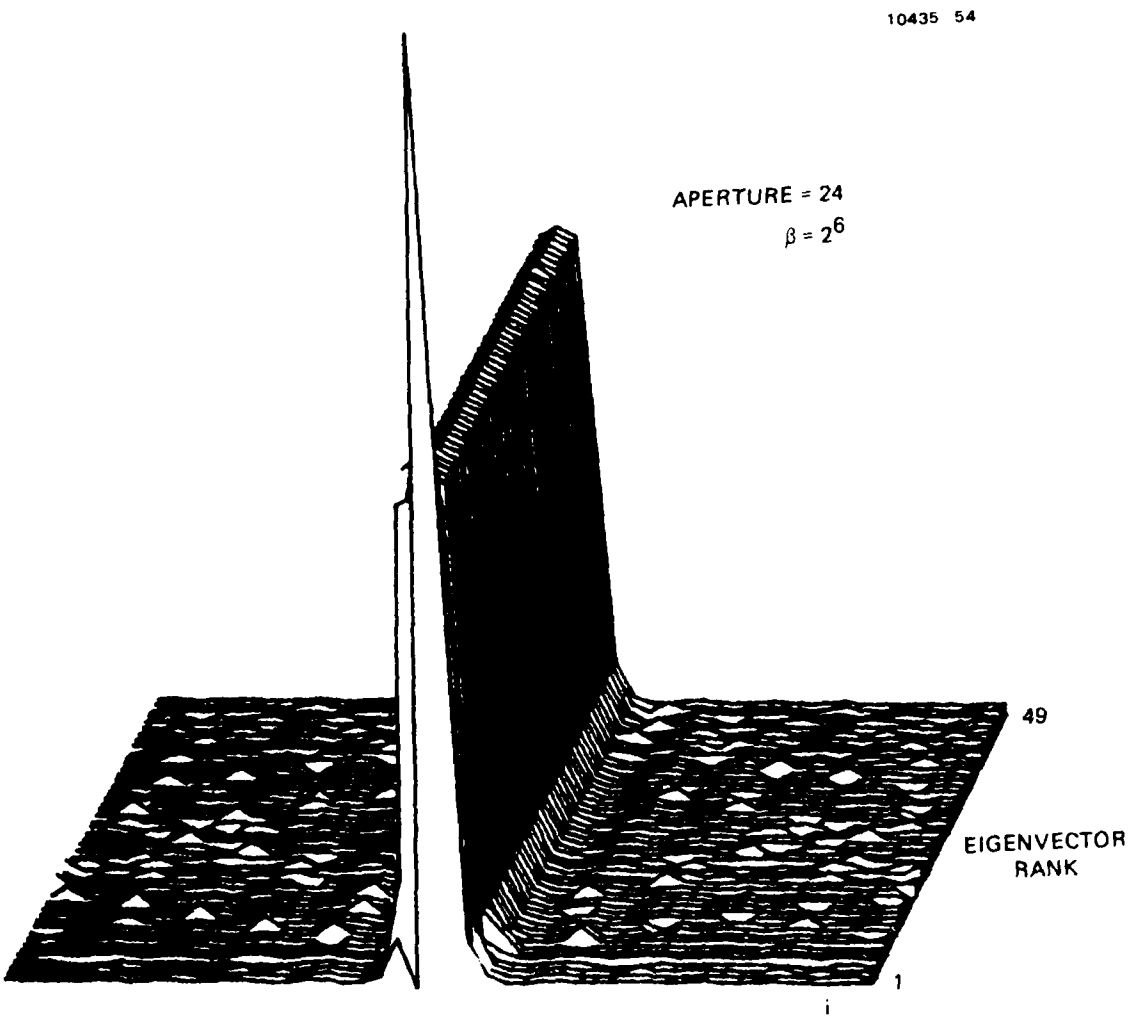


Figure 30(d). The ME estimates $\{p_i^{(o)}\}$ + [illegible] i for aperture 24, $\beta = 2^6$.

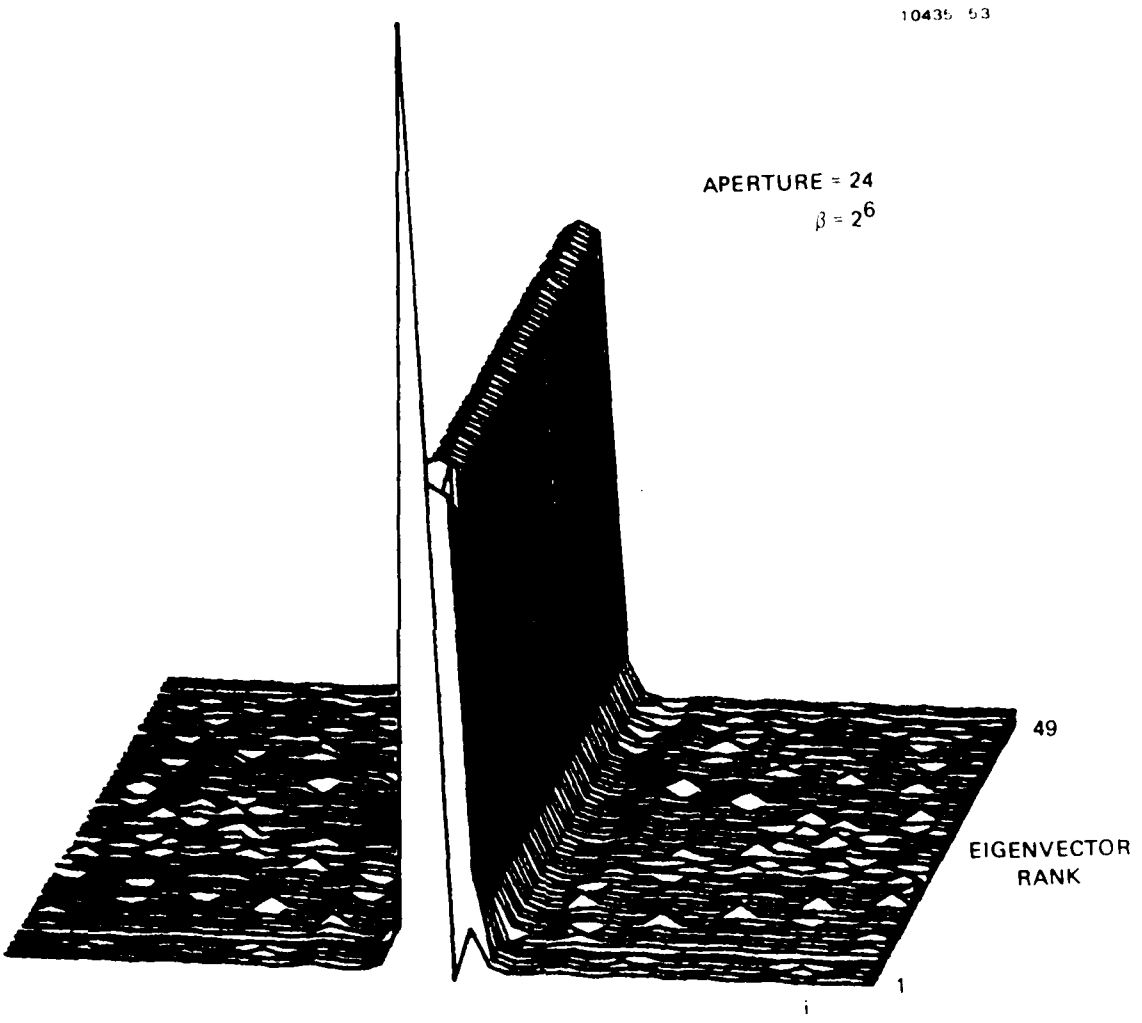


Figure 30(e). The ME estimates $p_i^{(o)}$ minus p_i for aperture 24, $\beta = 2^6$.

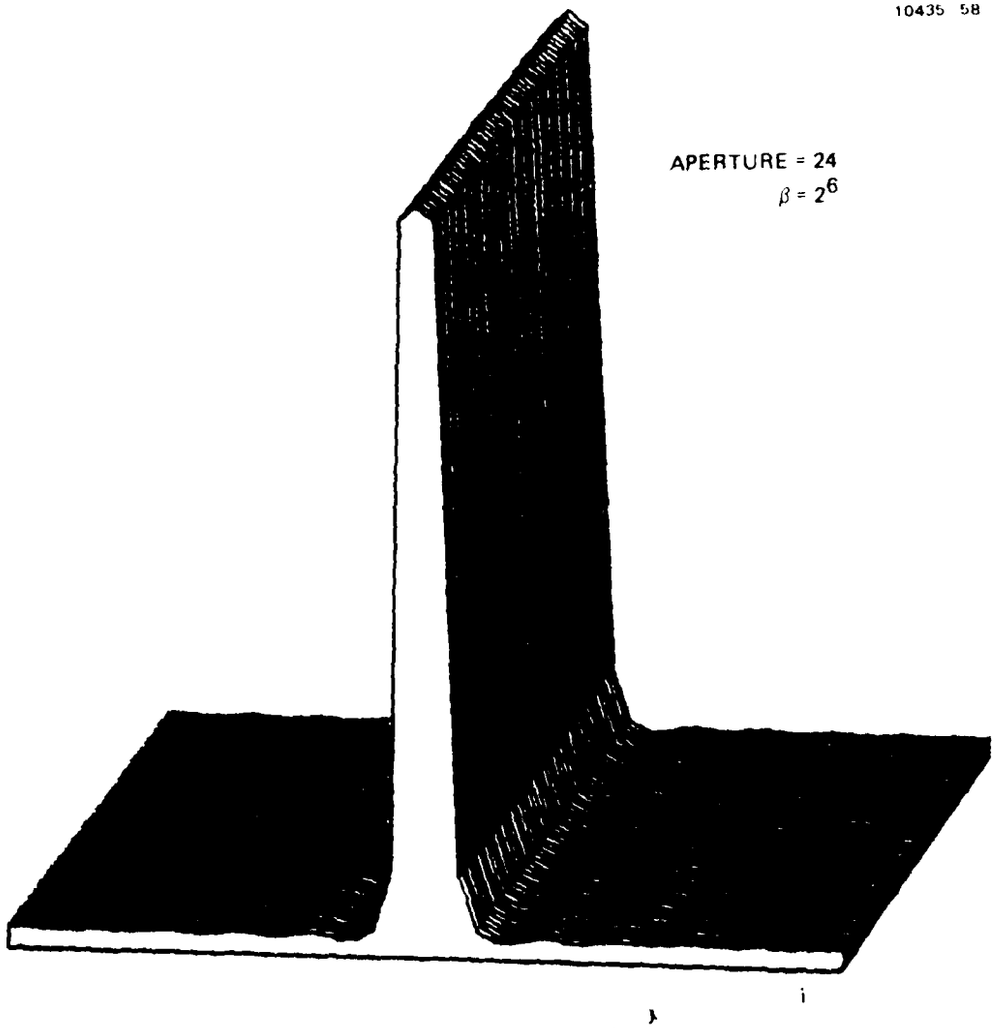


Figure 30(f). The ME estimate $\{p_i^{(o)}\}$ for aperture size 24, $\beta = 2^6$ replicated 49 times.

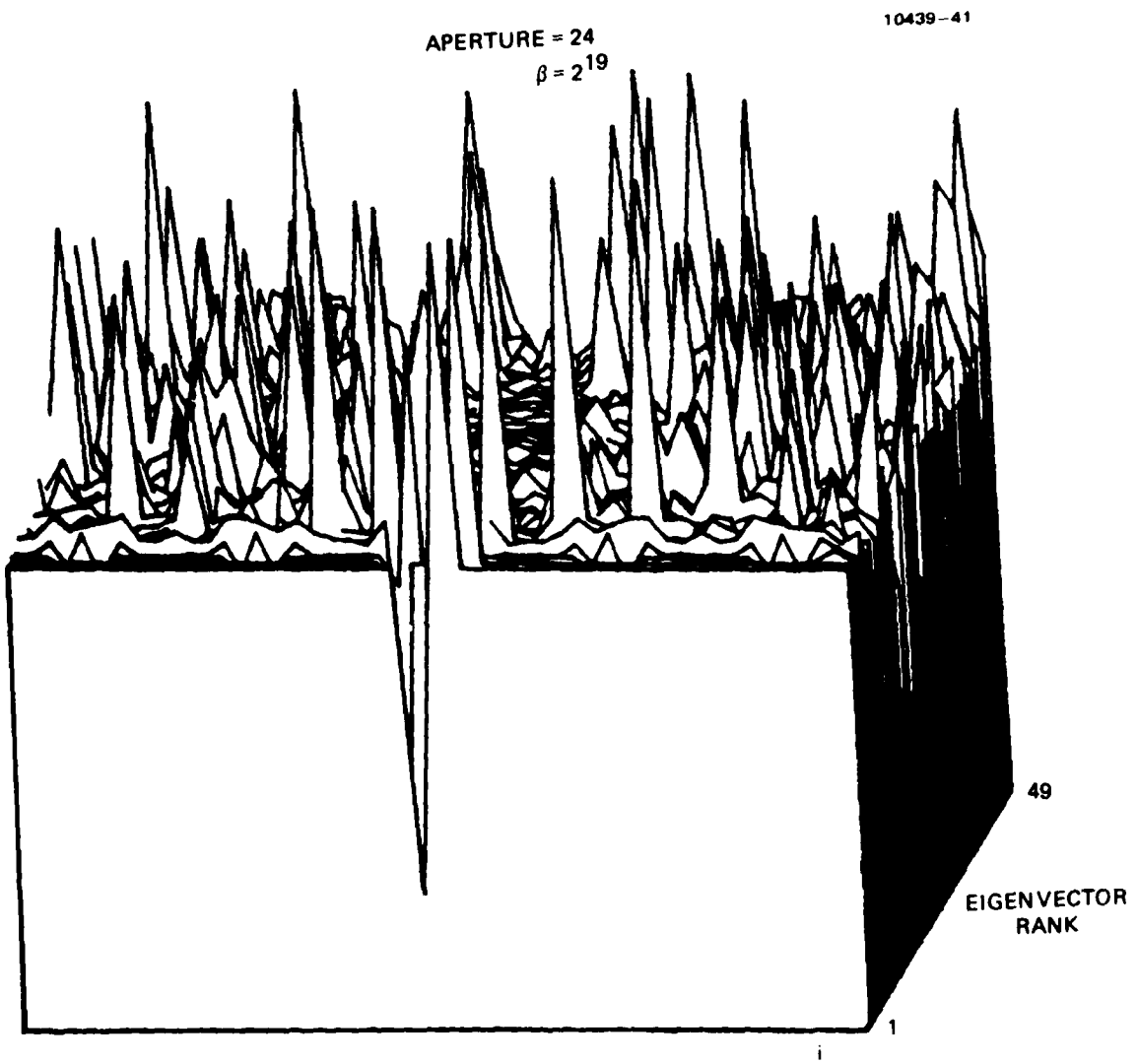


Figure 31(a). Orthonormal eigenvectors projected on p-space for aperture size 24 and $\beta = 2^{19}$, positive peaks.

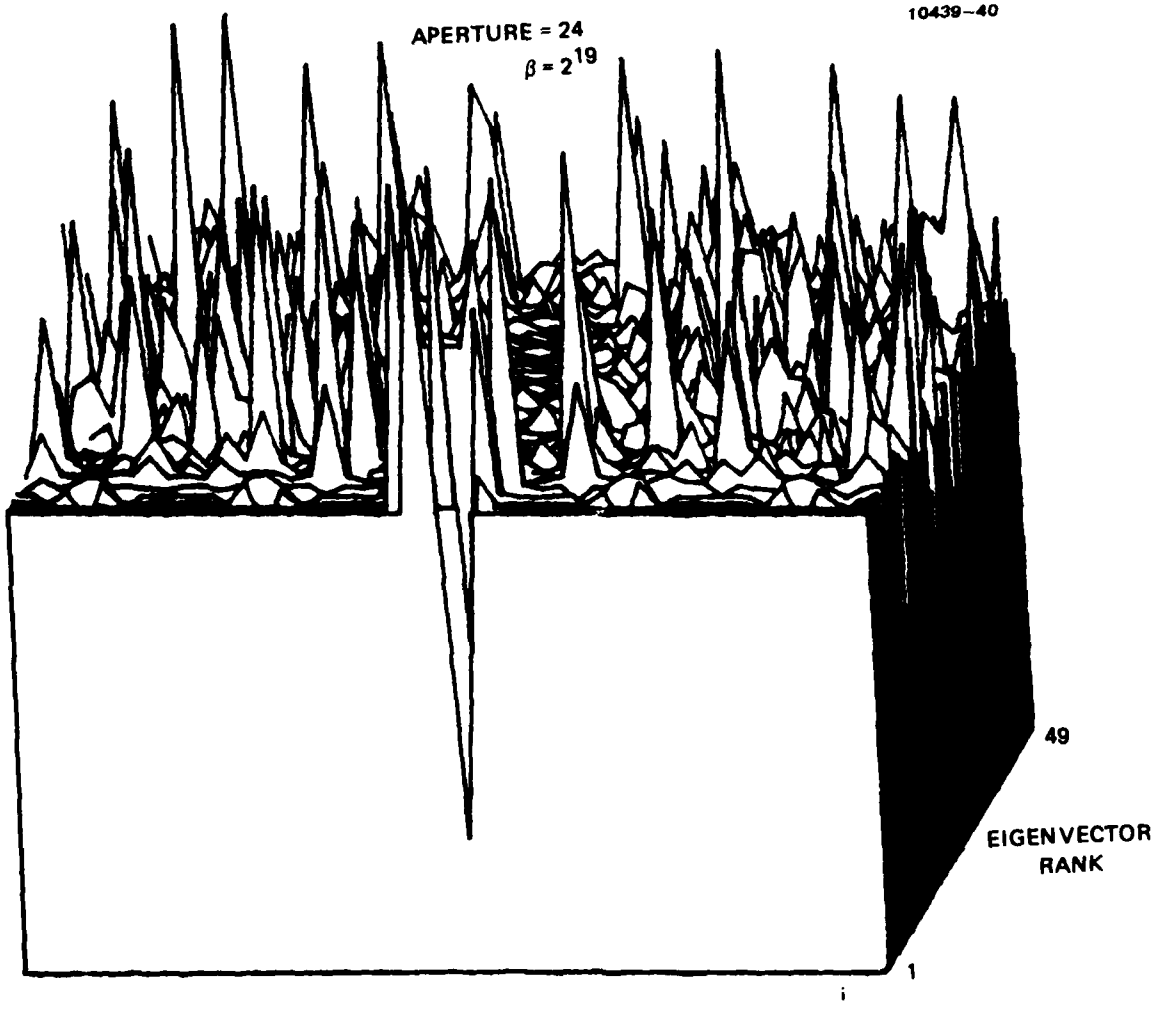


Figure 31(b). Orthonormal eigenvectors projected on p-space for aperture size 24 and $\beta = 2^{19}$, negative peaks.

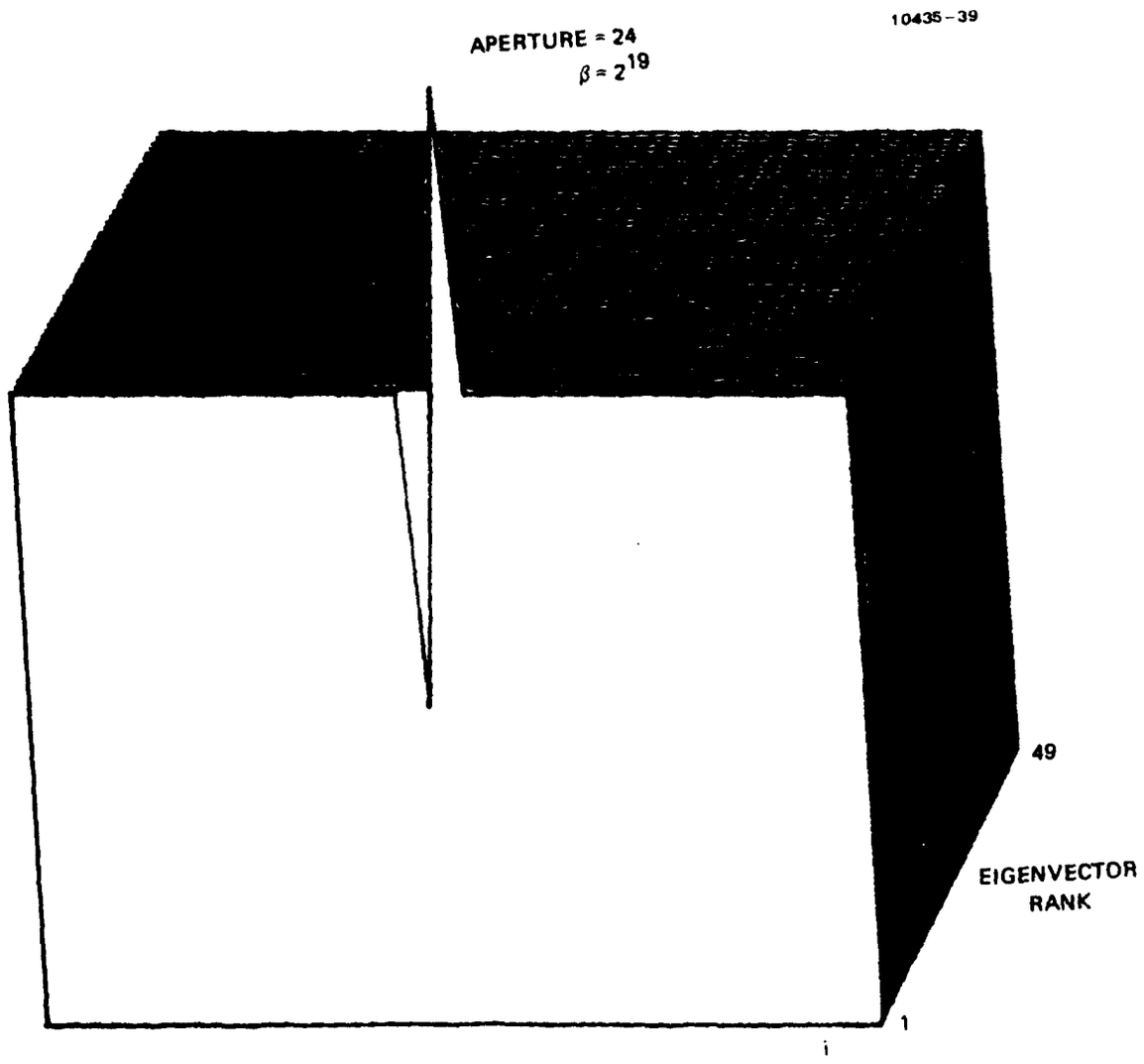


Figure 31(c). Eigenvectors rescaled and projected to p-space as $\sqrt{2(\sigma)}$ for aperture 24, $\beta = 2^{19}$.

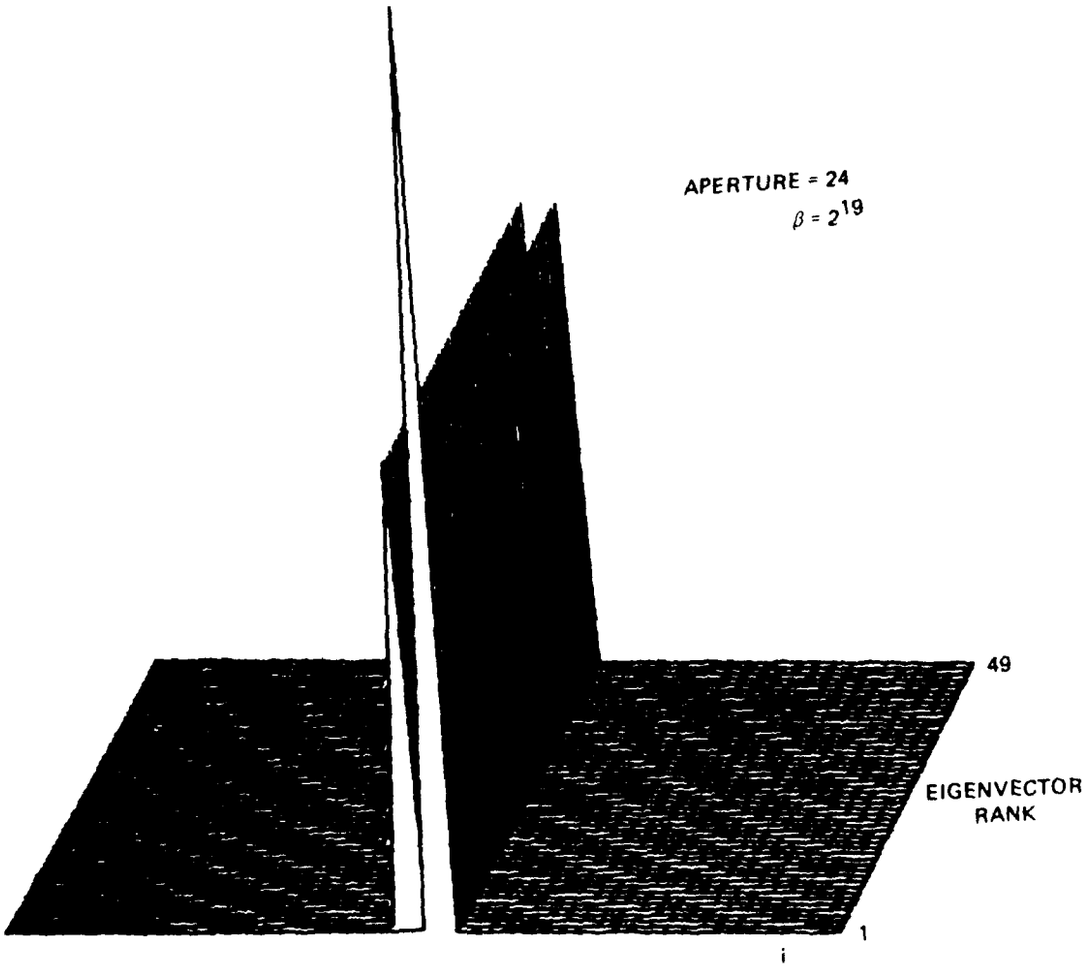


Figure 31(d). The ME estimate $\{p_i^{(o)}\}$ plus $\lambda 2\{\ell_i\}$ for aperture 24, $\beta = 2^{19}$.

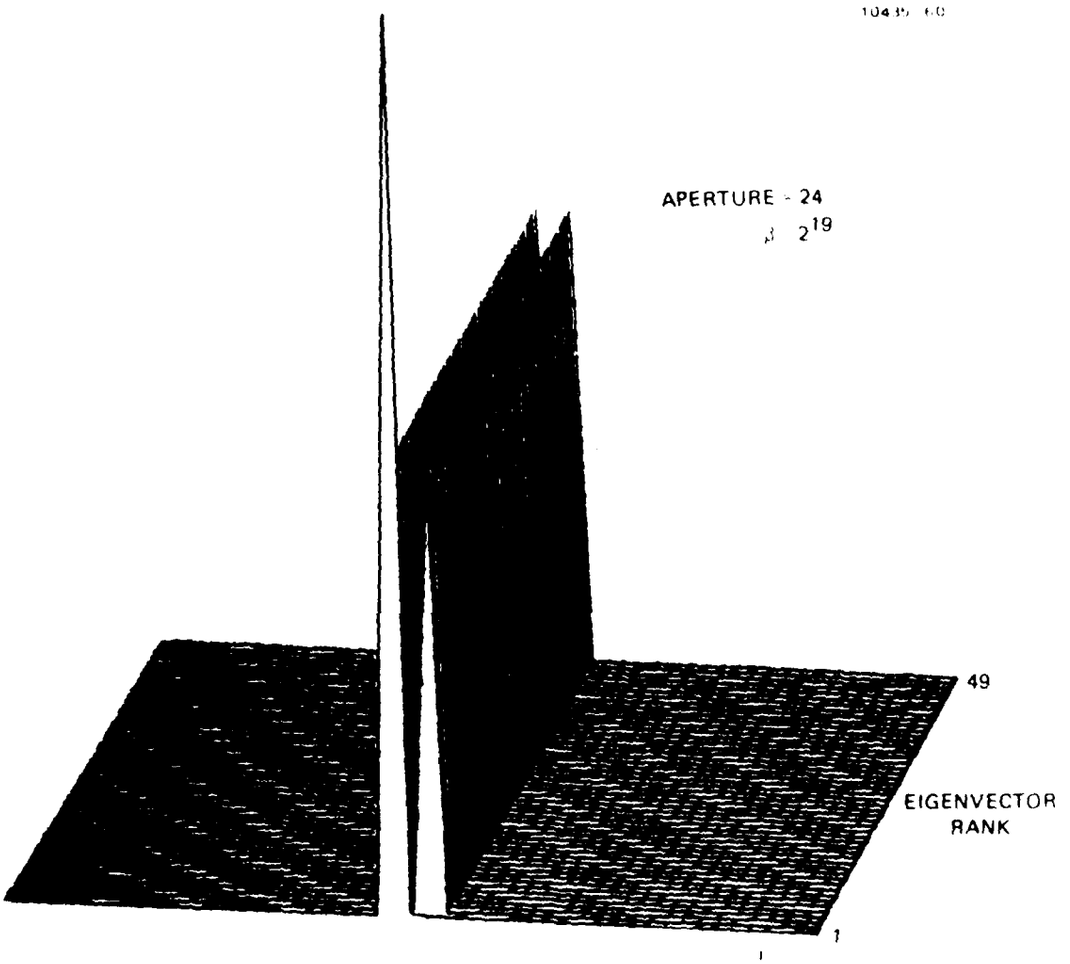


Figure 31(c). The ME estimate $\{p_i^{(\alpha)}\}$ minus $\{\ \}$ for aperture 24, $\alpha = 2^{19}$.

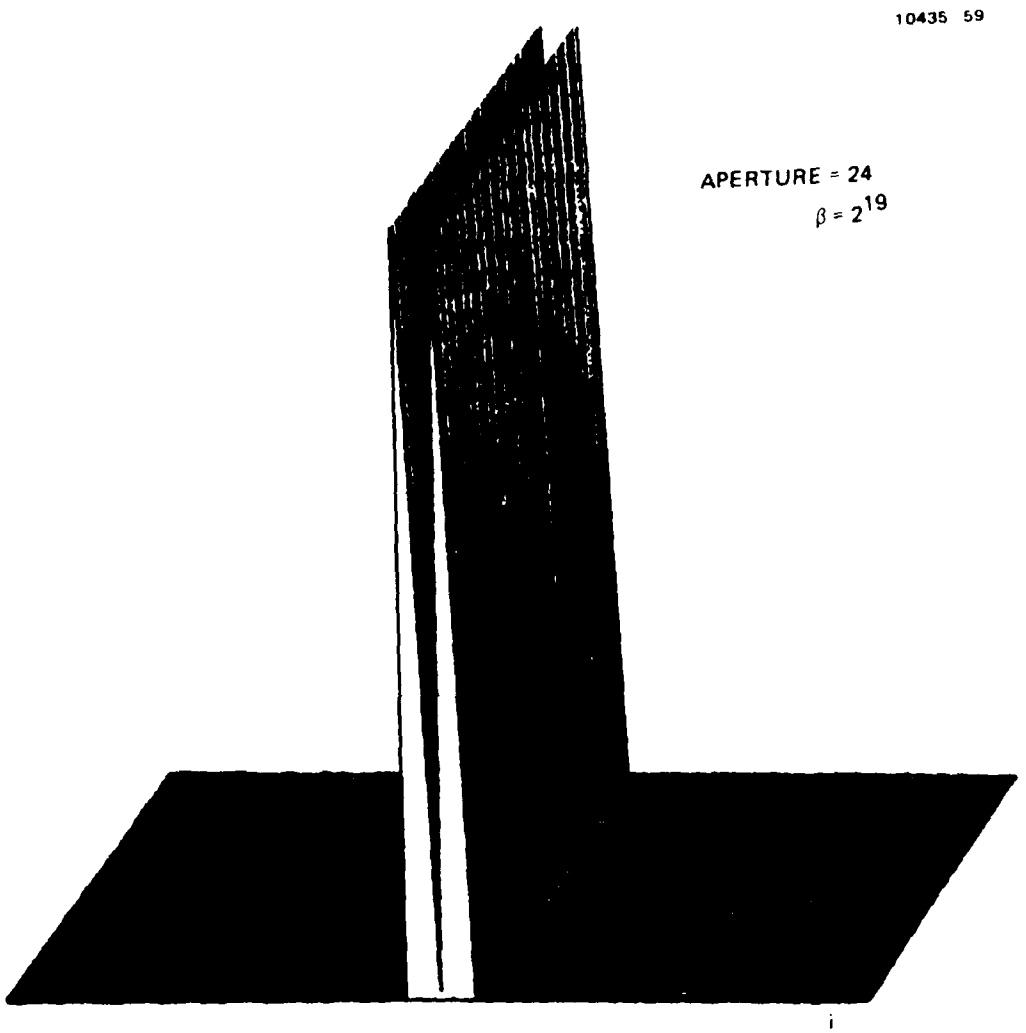


Figure 31(c). The ML estimate $\{\hat{p}_i^{(\infty)}\}$ for aperture size 24, 2^{19} replicated ∞ times.

given in Figures 32(a) and (b) for the standard and reverse view, respectively. In Figure 32(b) we see that the peaks are already resolved at $\beta = 2^0$. Even at this large aperture the total gamut of the σ_i is only a factor of two, and all the eigenvectors contribute to explain the variance as can be seen in Figure 33(c). The bounds of the confidence bands are shown in Figures 33(d) and (e), and the estimate itself $\{p_i^{(o)}\}$ is given in Figure 33(f).

The very last example, $\beta = 2^{18}$, aperture of 48, is shown in the series of Figure 34. In this extreme case the first σ is more than two thousand times larger than the second (see Table 2) and clearly accounts for almost all the variance. The hyperellipsoid, in this as in all the other high β cases, is a "hyperneedle." In Figure 34(c) the projected components of this vector can be seen to lie exactly on the peaks of the object estimate, which have values almost exactly reaching one half at the peaks (0.49866) and zero at the background (0.00006). The upper and lower confidence bands are shown in Figures 34(d) and (e).

It should not be assumed, by way of generalization, that the important eigenvectors will, in all cases, mimic the shape of the estimate at high β values as seen in this example. In another example, not illustrated here, with a one spike object, the important eigenvectors did cluster around the estimated spike, but had both symmetric and antisymmetric forms whose large components did not necessarily exactly coincide with the position of the object peak.

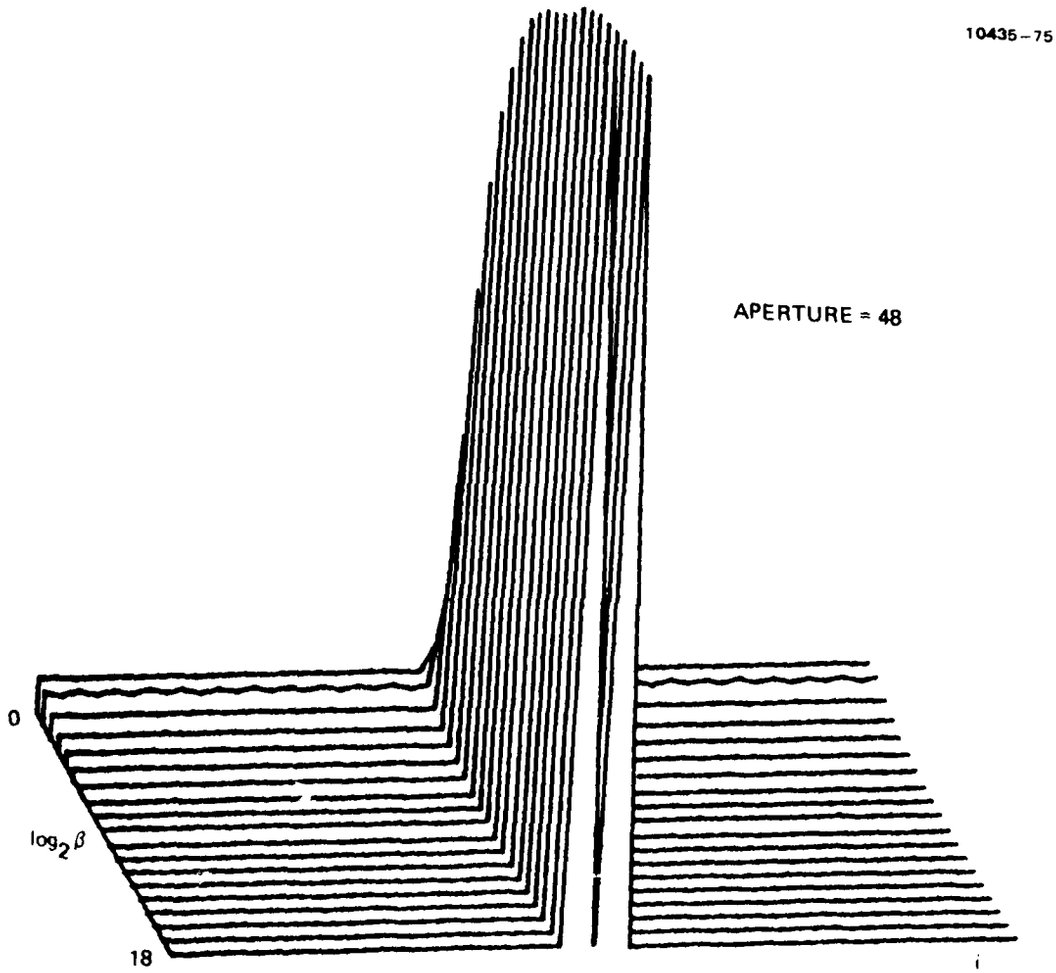


Figure 32(a). The ME estimates $\{p_i^{(o)}\}$, $1 \le i \le 49$, for aperture size 48, and for $2^0 \le \beta \le 2^{18}$. Standard view.

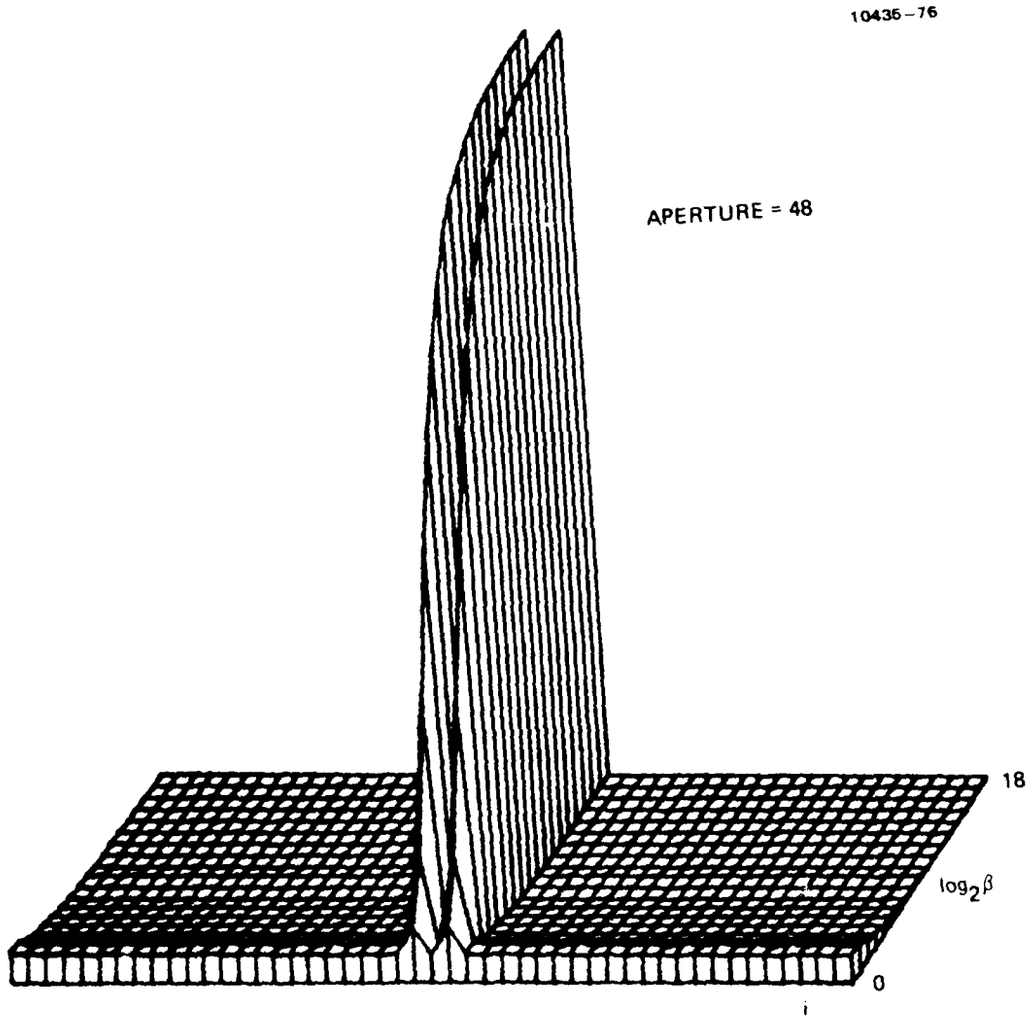


Figure 32(b). The ME estimates $\{p_i^{(o)}\}$, $1 \le i \le 49$, for aperture size 48 and for $2^{-18} \le \beta \le 2^{18}$. Reverse view.

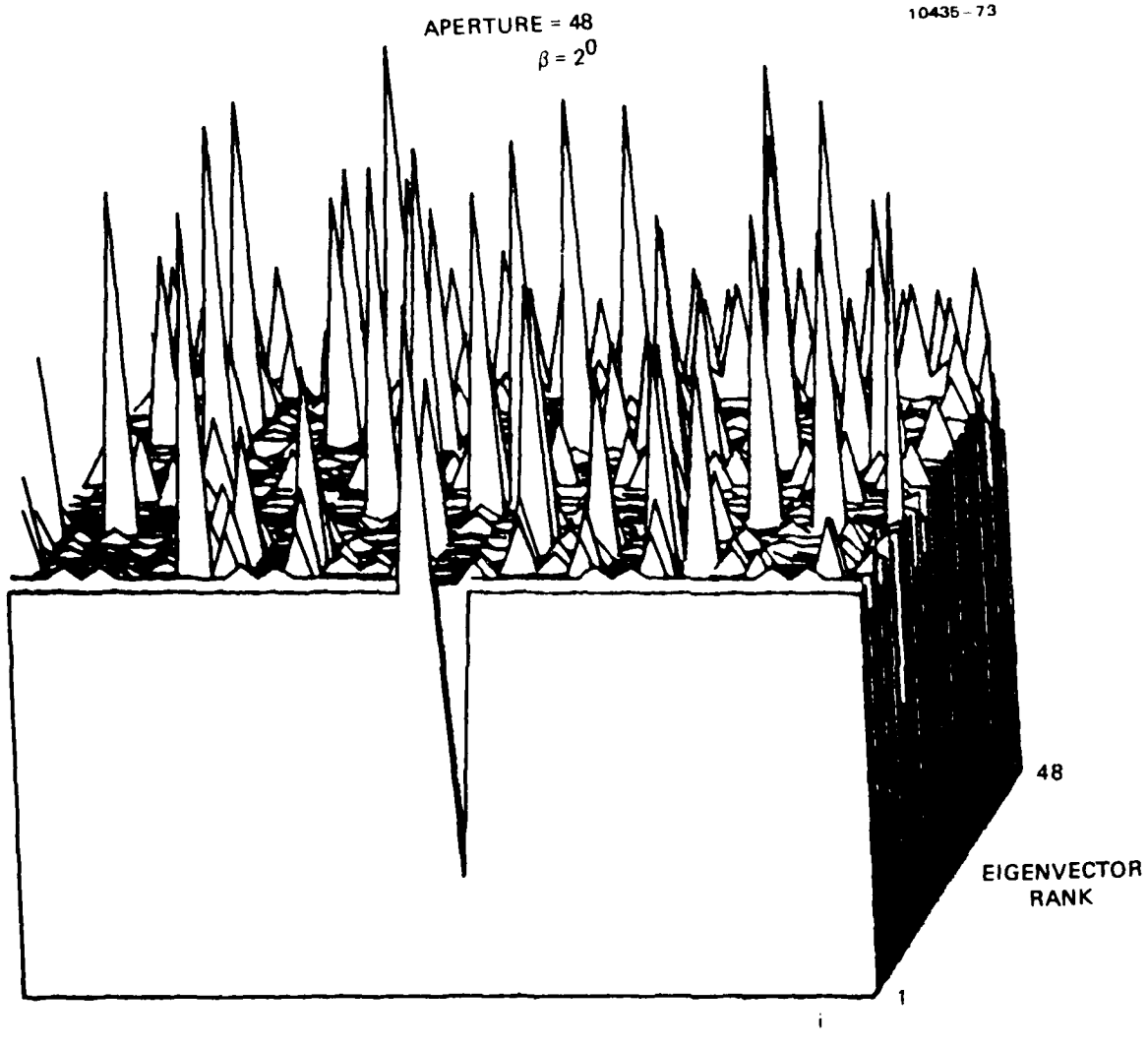


Figure 33(a). Orthonormal eigenvectors projected on p-space for aperture size 48 and $\beta = 2^0$. Positive peaks.

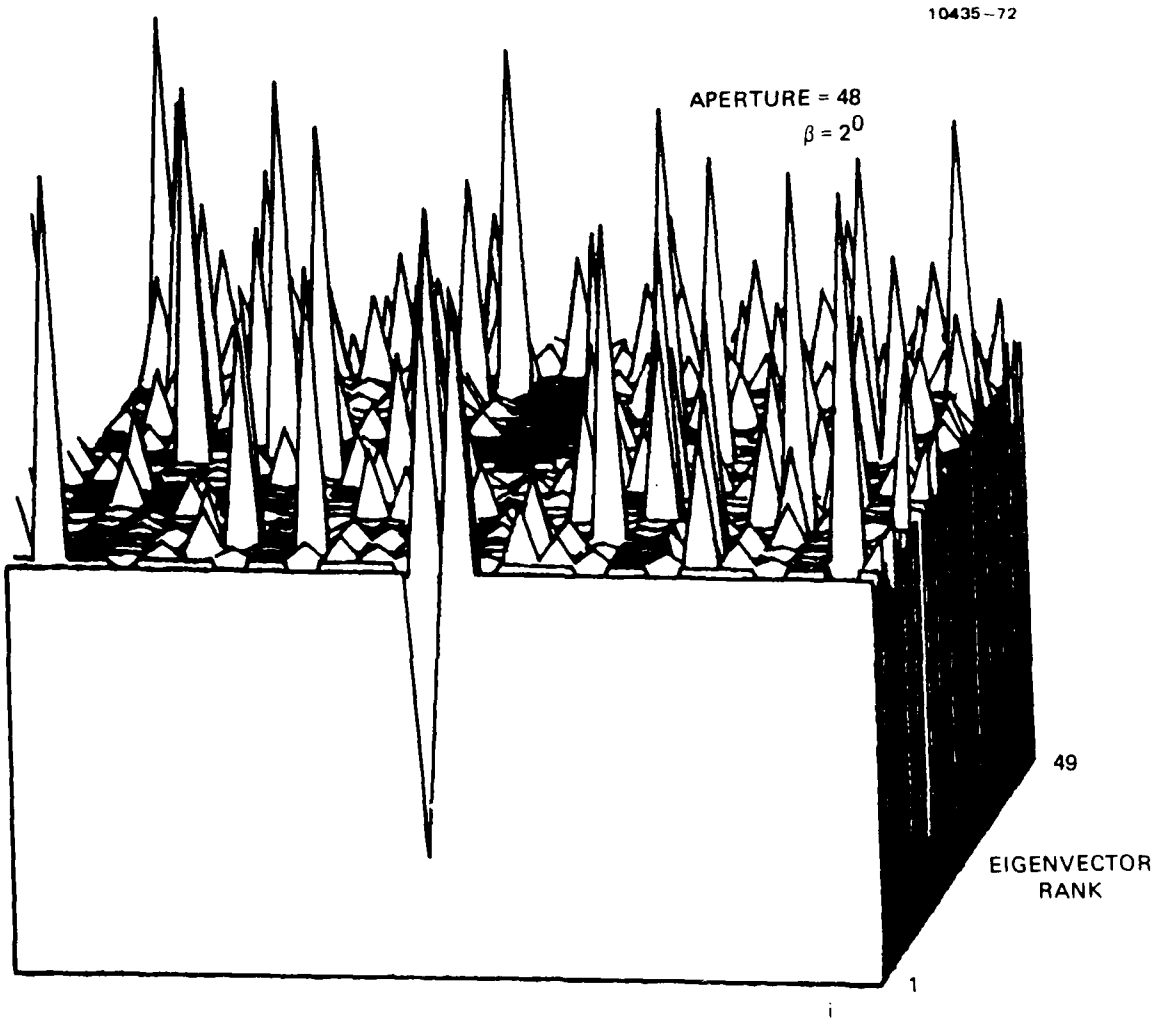


Figure 33(b). Orthonormal eigenvectors projected on p-space for aperture size 48 and $\beta = 2^0$. Negative peaks.

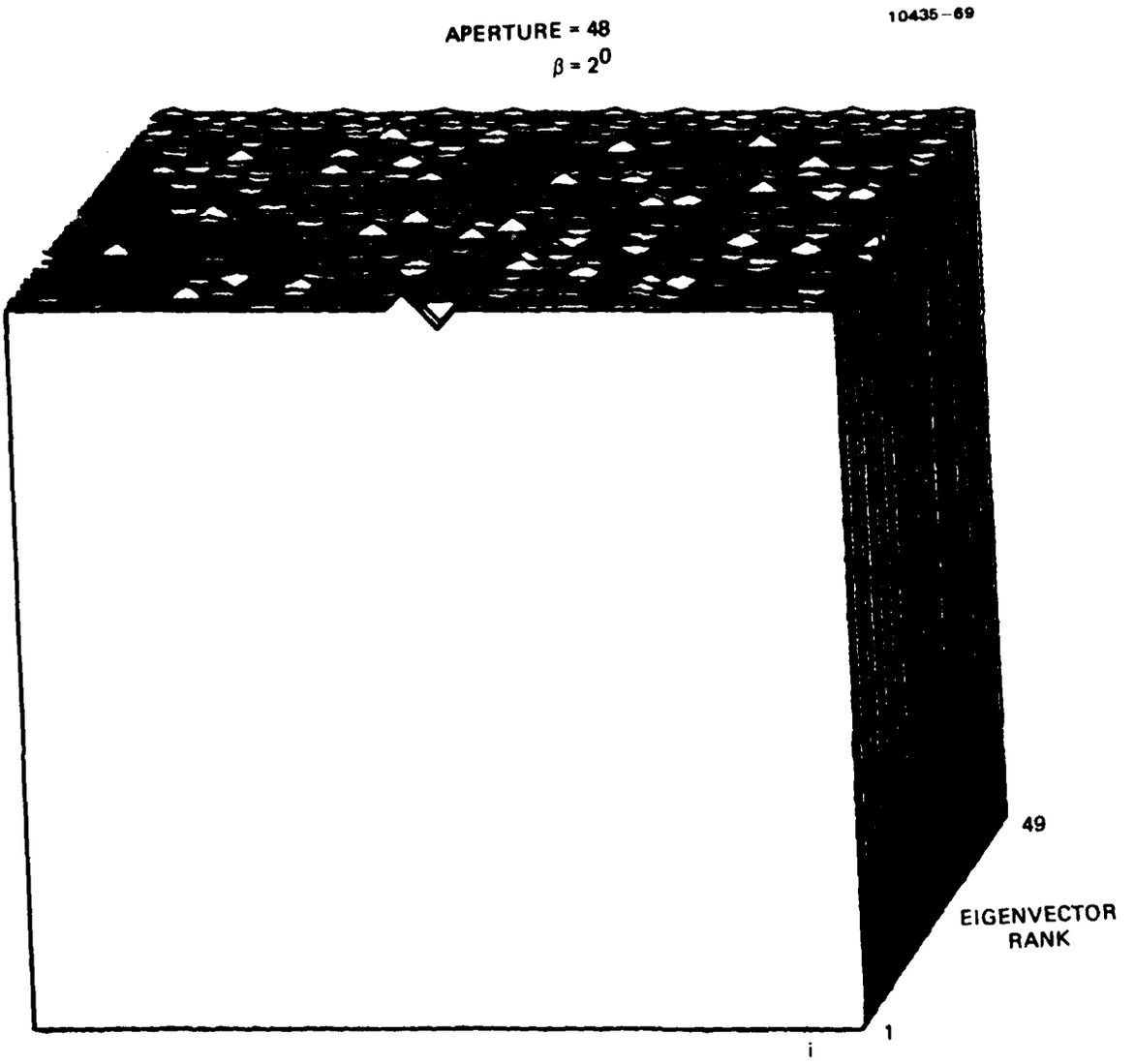


Figure 33(c). Eigenvectors rescaled and projected to p-space as 1.03 $\{\sigma_i\}$ for aperture 48, $\beta = 2^o$.

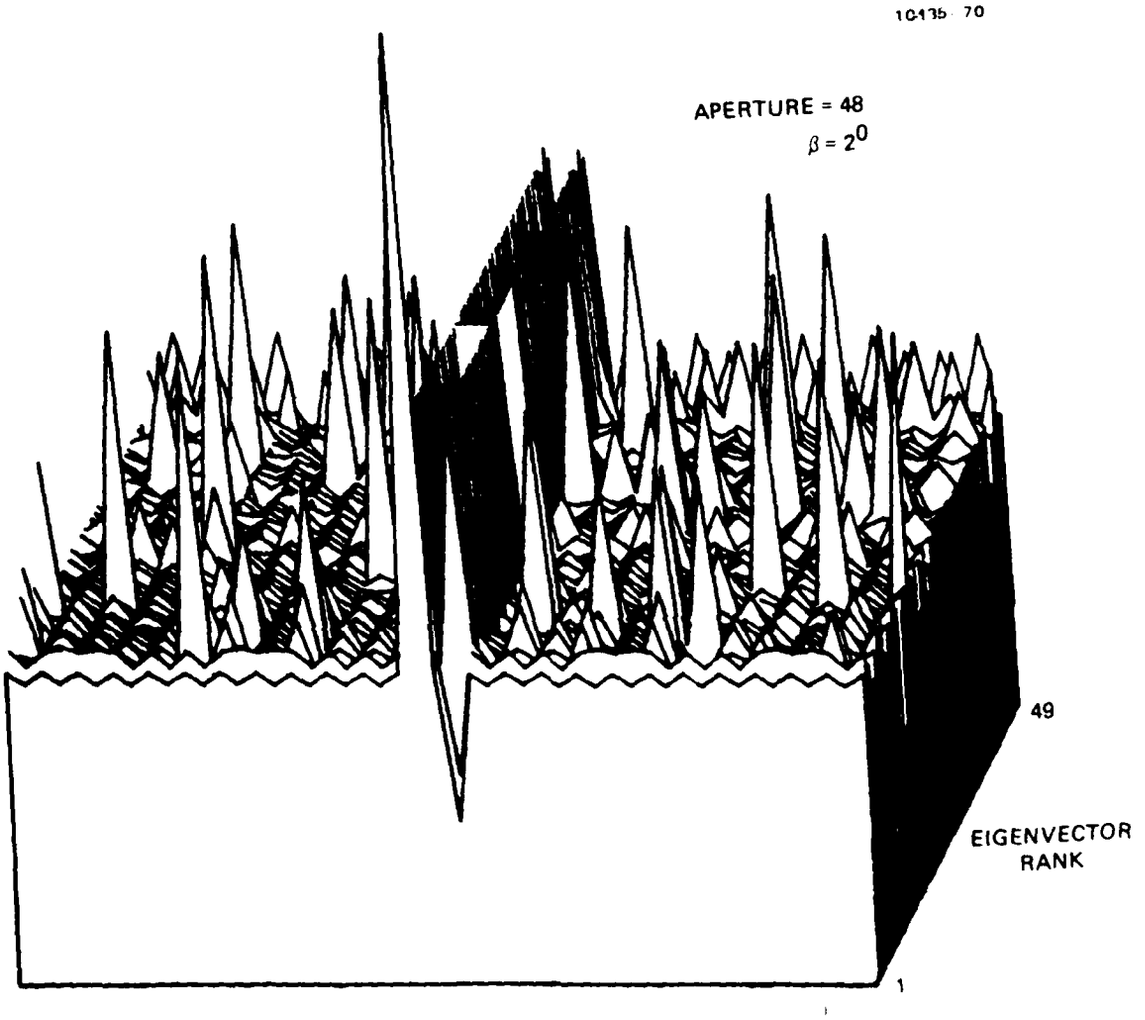


Figure 33(d). The ME estimates $\{P_i^{(o)}\}$ plotted as $\{\ \}$ for aperture 48, $\beta = 2^0$.

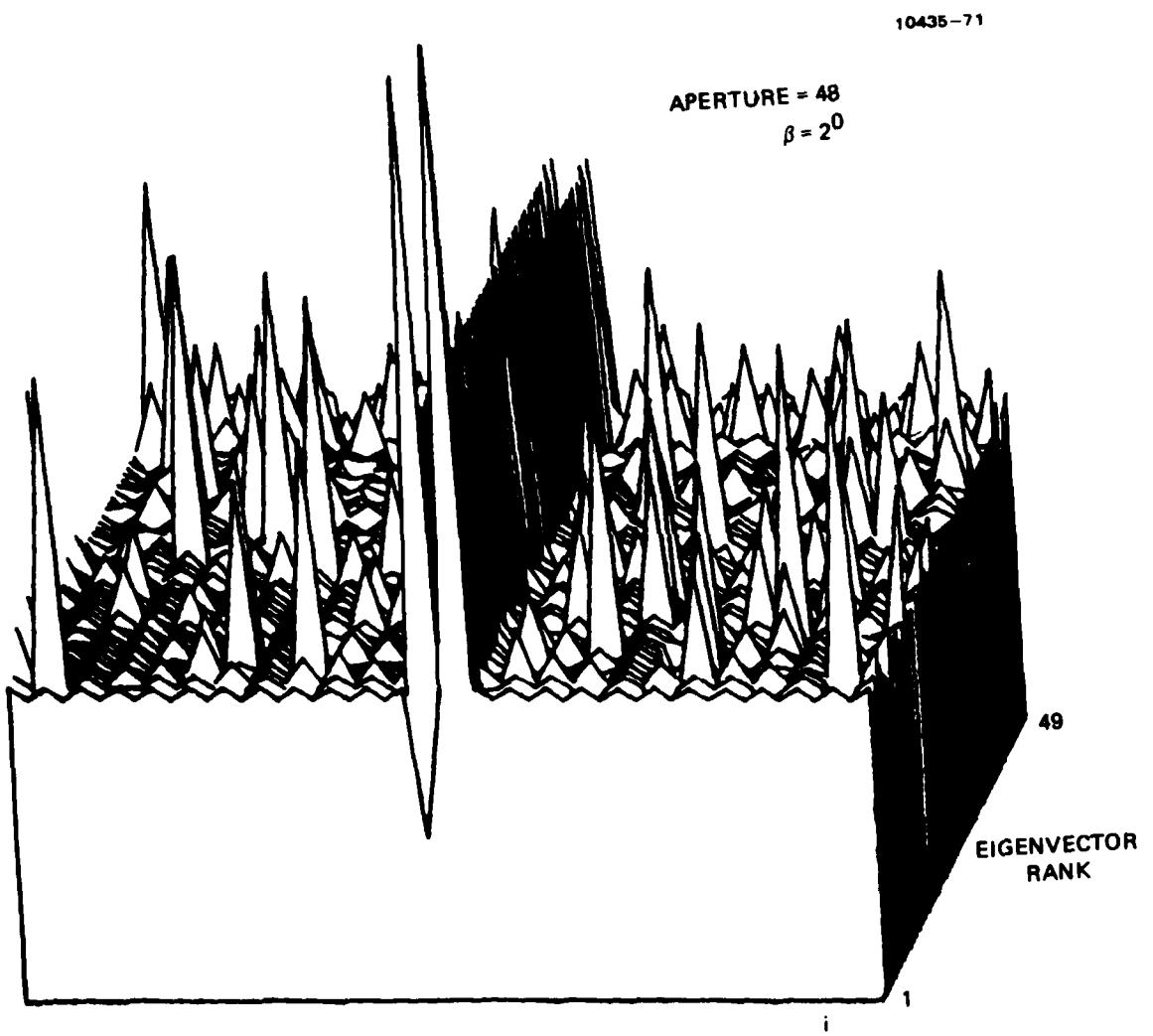


Figure 33(e). The ME estimates $\{p_i^{(o)}\}$ minus $1.03 \cdot \{\sigma_i\}$ for aperture 48, $\beta = 2^o$.

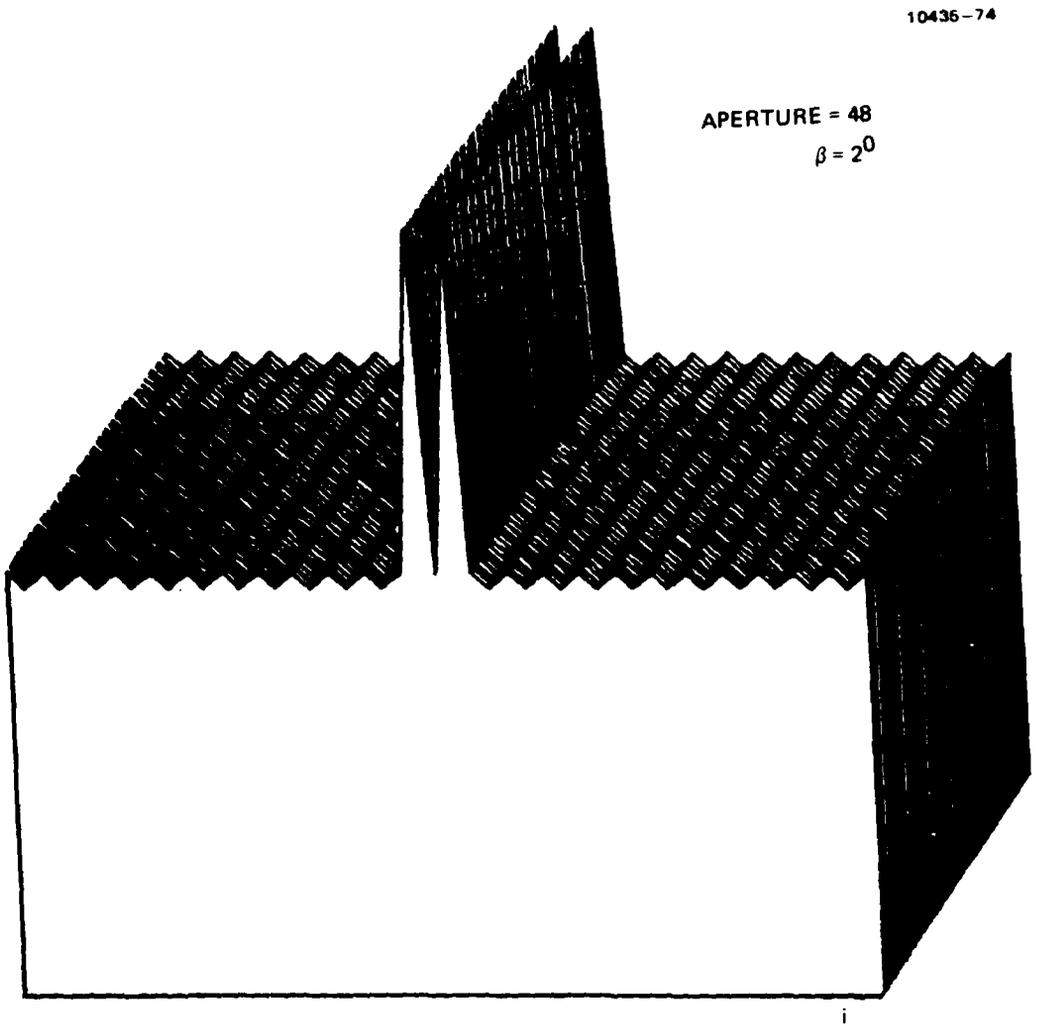


Figure 33(f). The ME estimates $\{p_i^{(o)}\}$ for aperture size 48 and for $\beta = 2^o$ replicated 49 times.

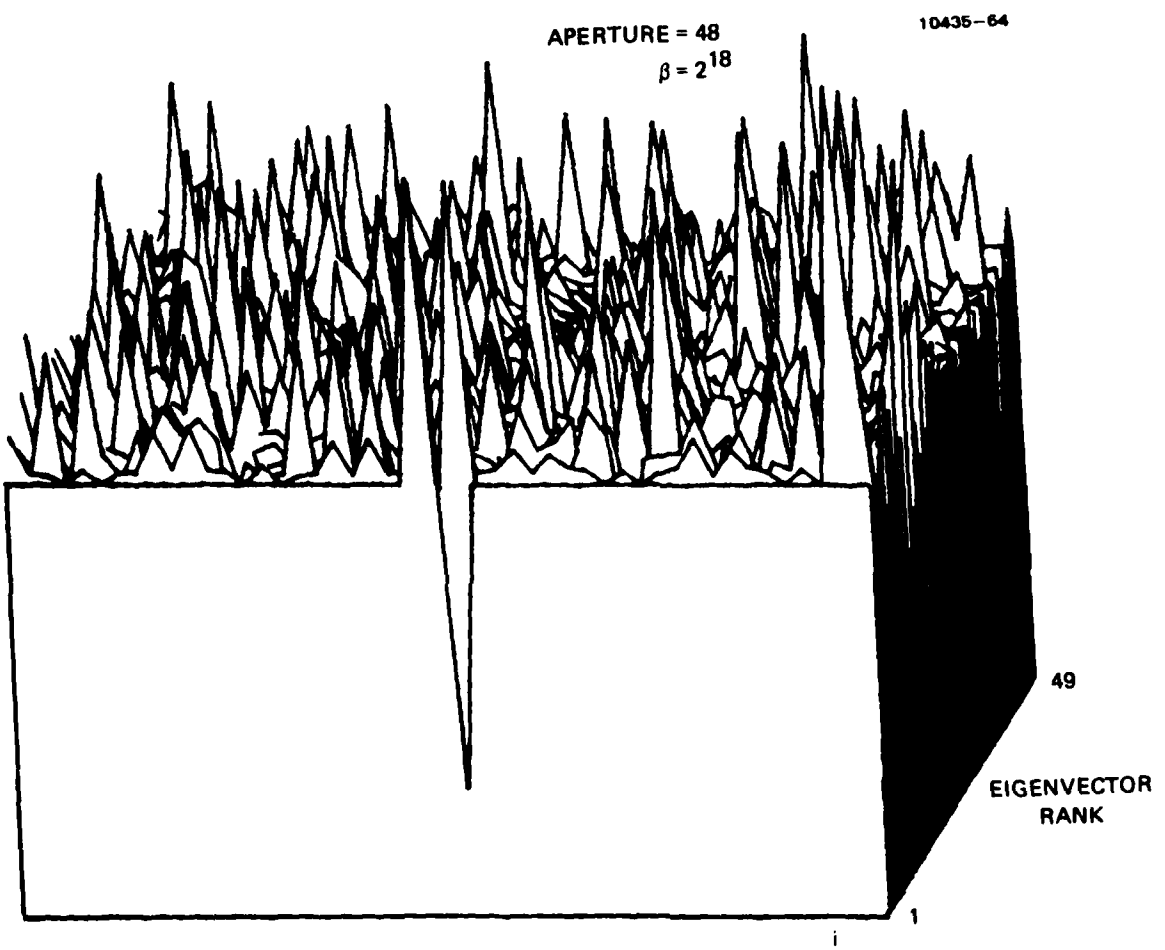


Figure 34(a). Orthonormal eigenvectors projected on p-space for aperture size 48 and $\beta = 2^{o}$. Positive peaks.

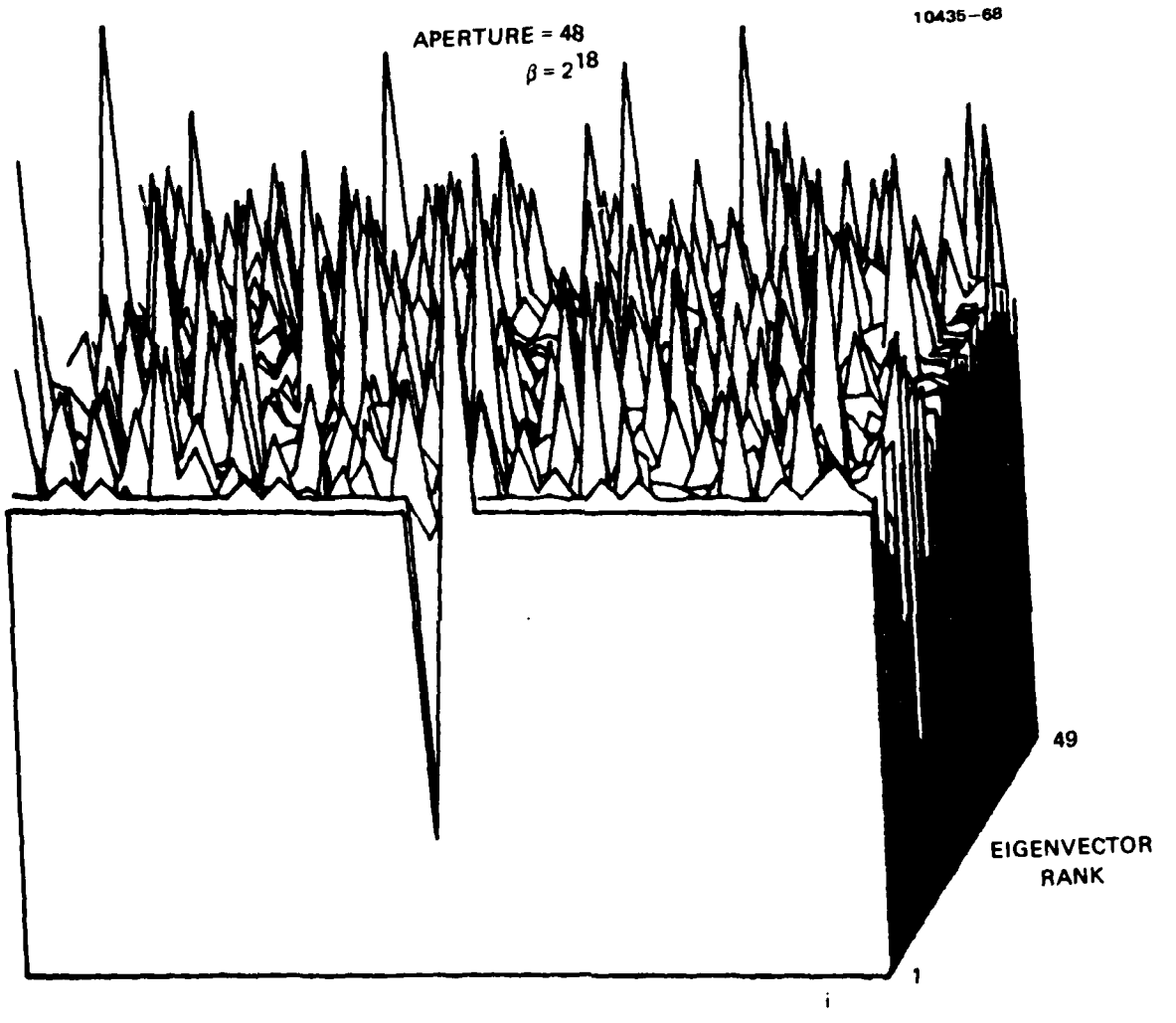


Figure 34(b). Orthonormal eigenvectors projected on p-space for aperture size 48 and $\beta = 2^{18}$. Negative peaks.

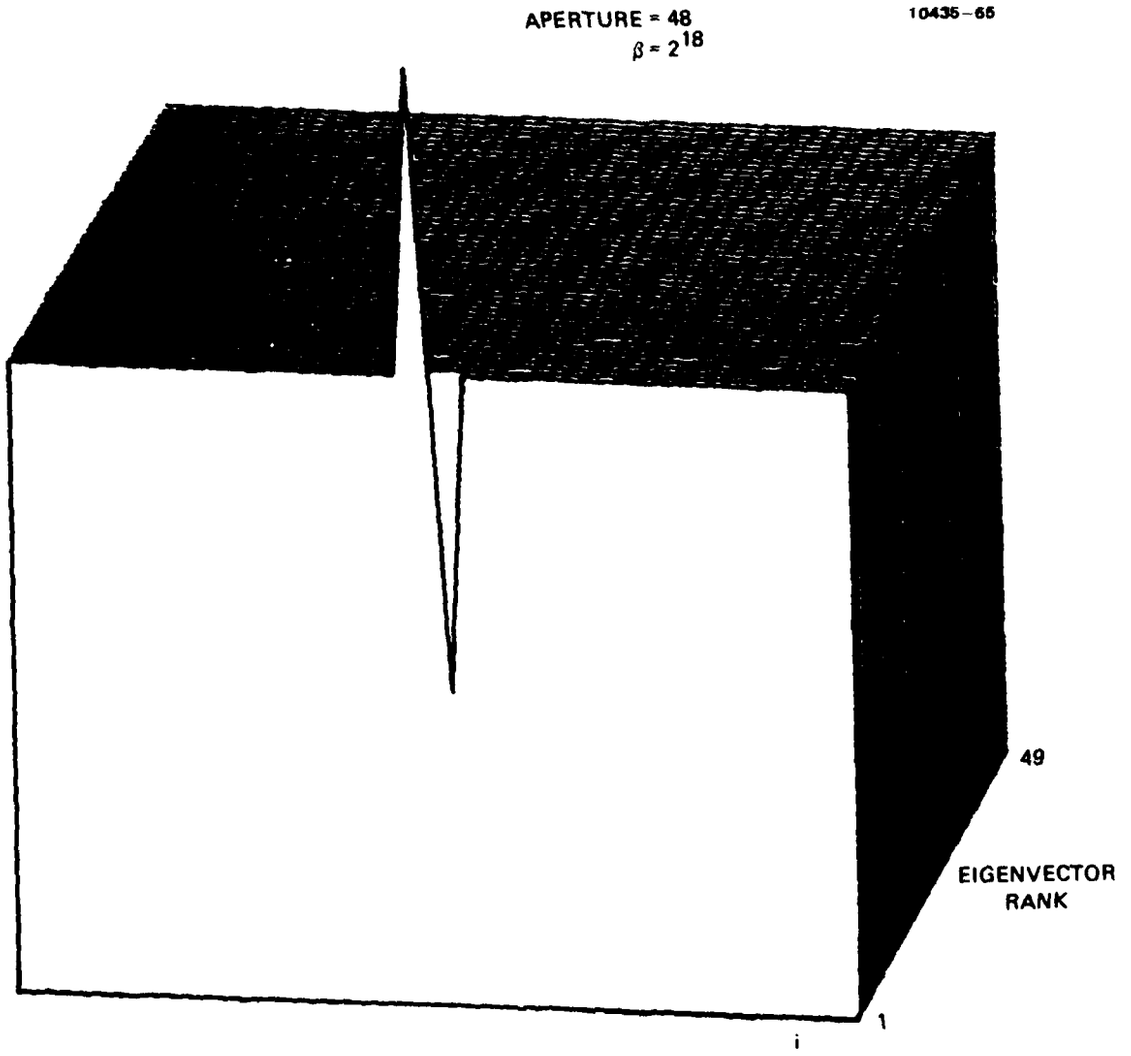


Figure 34(c). Eigenvectors rescaled and projected to p-space as $\sqrt{2}\{\sigma_i\}$ for aperture 48, $\beta = 2^{18}$.

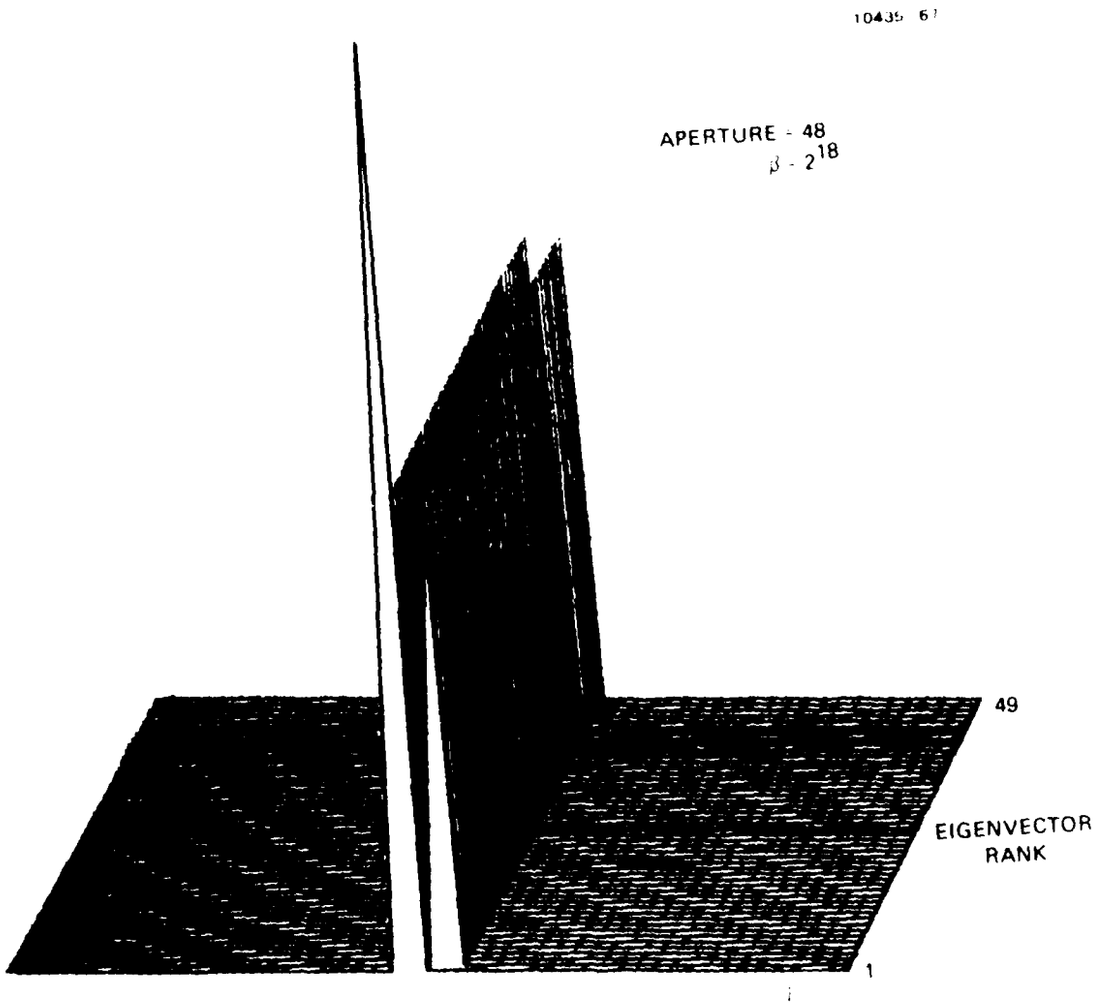
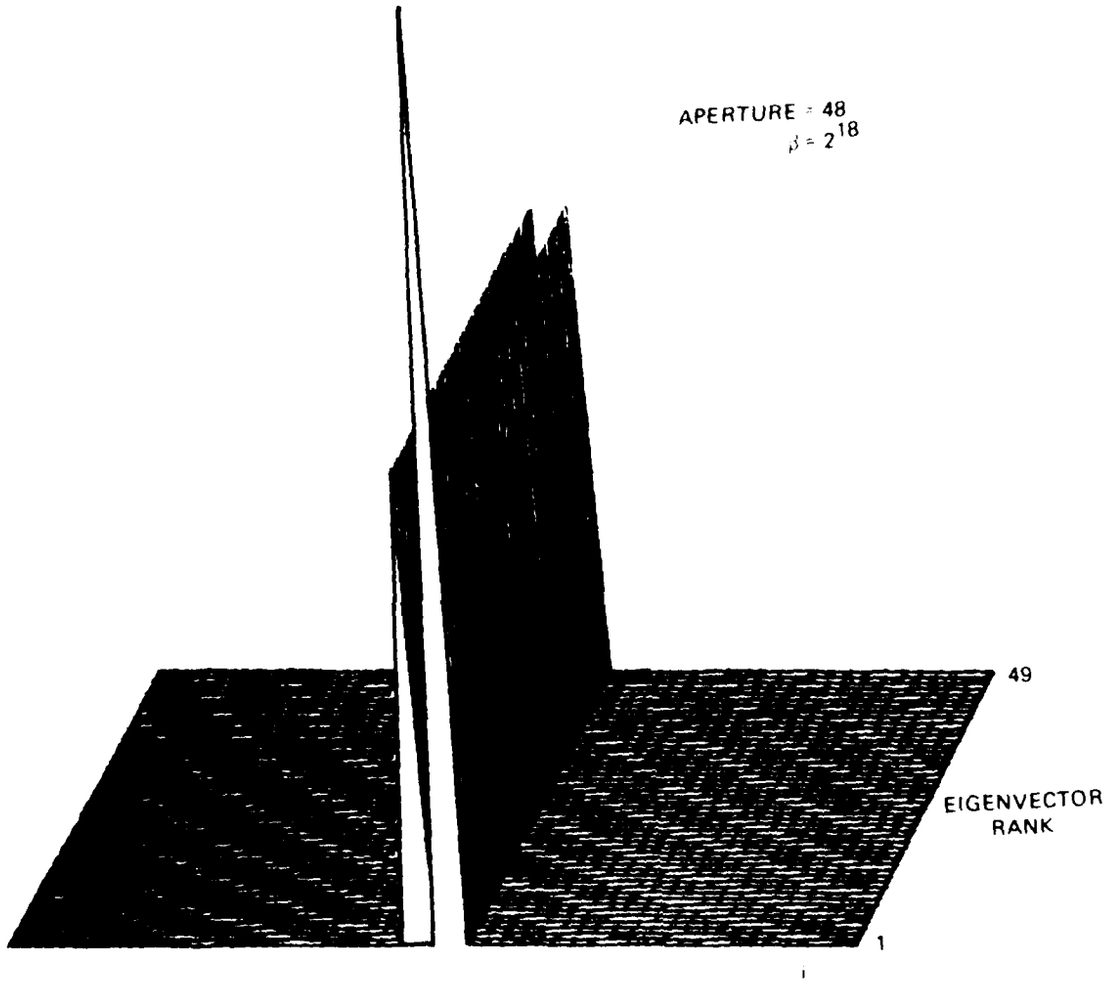


Figure 34(d). The ME estimates $\{p_i^{(o)}\}$ plus $\{\cdot\}$ for aperture 48, $\beta = 2^{18}$.


APERTURE = 48
$\beta = 2^{18}$
49
EIGENVECTOR
RANK
1

[illegible]

SECTION 6

SUMMARY OF RESULTS

The work on the simulation of imaging with thinned, sparse, random arrays, done in the first two years of this contract, was summarized in the Annual Report, June 1979. In this section, we summarize the final year's work on the ME method. We present a general view of maximum entropy (ME) as a method to choose the most probable object estimate from the set of all the possible objects consistent with the degraded measurements. The probabilities are assigned in accordance with the physical statistics of the problem at hand. The entropy is the logarithm of the probability. We show that the two contending forms of entropy found in the current literature, the "logB" and "BlogB" forms, are special cases of a more general entropy based on quantum statistics. Our particular formulation of the ME method can be expressed as an analogy to the thermodynamical method of the minimization of the free energy. This analogy allows a natural way to introduce the matter of fluctuations or noise into the method. The problem of a complete theory of the fluctuations still remains, awaiting a theory of the analogy to the specific heat.

Several examples of ME estimation, in one and two dimensions and for varying values of noise parameter, were calculated and the results are presented graphically. The limiting effects of noise on the possibilities of superresolution are described.

A general method for assessing the degree of confidence in the multidimensional ME estimates was developed. The ME formulation we developed provides a multidimensional probability function for all possible potential object estimates consistent with the measured-image signal data. The most probable one is picked as the estimate. We expand this distribution function near its maximum for simplicity, and approximate it as a multivariate Gaussian distribution. By principal axis transformations, we derive the variances and then project them

back to the original space of the estimates as confidence bands. A chi-square test in many dimensions is used as the basis of the confidence estimate. A particular example was studied in some detail. Two delta-function spikes on a support of dimension 49, separated by one space, was chosen as the unknown test object. Various images were derived by choosing progressively smaller apertures to test the method as a function of superresolution and fluctuation or noise temperature. Many variations were presented spanning a range of inverse fluctuation or noise temperature from 2^0 to 2^{19} and potential superresolution factors of from one through eight.

SECTION 7

PERSONNEL

Bernard H. Soffer was the program manager. The principal investigators on the program described in this report were Bernard H. Soffer and Ryoichi Kikuchi. Karen Olin made helpful contributions in computer graphics.

Contributors to the earlier phase of the work are listed in the Annual Report, June 1979, F49620-77-C-0052.

SECTION 8

PAPERS RESULTING FROM AFOSR SUPPORT

Comments on "Spectral Estimation: An Impossibility?" B.H. Soffer and Ryoichi Kikuchi, Proc. IEEE 67, 1672 (1979).

In preparation are: "Maximum Entropy Image Restoration II and III, to be submitted to JOSA.

REFERENCES

1. S.J. Wernecki and L.R. D'Addario, "Maximum Entropy Image Reconstruction," IEEE Trans. Computers C-26, 351 (1977); S.J. Wernecke, "Two Dimensional Maximum Entropy Reconstruction of Radio Brightness," Radio Sci 12, 831 (1977).

2. R. Kikuchi, "Superposition Approximation and Natural Iteration Calculation in Cluster-Variation Method," J. Chem. Phys. 60, 1071 (1974).

3. A. Ralston, "A First Course in Numerical Analysis" (McGraw-Hill, NY, 1965).

4. S.F. Gull and G.J. Daniell, Nature 272, 686 (1978).

APPENDIX A

Maximum entropy image restoration. I. The entropy expression

Ryoichi Kikuchi and B. H. Soffer
Hughes Research Laboratories, Malibu, California 90265
(Received 3 July 1976; revision received 11 August 1977)

The two entropy expressions, $\log B$ and $-B\log B$ (where B is the local brightness of the object or its spatial spectral power) used in maximum entropy (ME) image restoration, are derived as limiting cases of a general entropy formula. The brightness B is represented by the n photons emitted from a small unit area of the object and imaged in the receiver. These n photons can be distributed over z degrees of freedom in $q(n,z)$ different ways calculated by the Bose-Einstein statistics. The entropy to be maximized is interpreted, as in the original definition of entropy by Boltzmann and Planck, as $\log q(n,z)$. This entropy expression reduces to $\log B$ and $-B\log B$ in the limits of $n \gg z > 1$ and $n \ll z$, respectively. When n is interpreted as an average $\bar{n}$ over an ensemble, the above two criteria remain the same (with n replaced by $\bar{n}$), and in addition for the $z = 1$ case the $\log B$ expression, used in ME spectral power estimation, is derived for $\bar{n} \gg z = 1$.

I. INTRODUCTION

The ideal of maximum entropy (ME) restoration has appeared in various forms in the technical literature spanning a gamut of disciplines from seismic spectral power estimation and prediction to astronomical image restoration. The subject was recently reviewed,[1] and an extensive bibliography was given.[2]

The concept, as applied to the image-restoration problem, can be viewed as an attempt to find the radiant spatial power pattern of the object by maximizing the entropy of the pattern subject to the image-data constraints. However, the method has a weakness in its very foundation. There are two schools of thought which differ as to the form of the entropy to be maximized. For convenience, we refer, in the remainder of this paper, to these two as the "$\log B$" and the "$-B\log B$" schools. (Although the B is defined more fully below, B may be interpreted, for purposes of this introduction, as the brightness of the object or as its spatial spectral power.) In his pioneering work in spectral power estimation for geophysical applications, Burg[3] used the $\log B$ expression for the entropy, following directly from Shannon's[4] work. In the image-restoration field, Ponsonby[5] and Wernecke and D'Addario[2] also used the $\log B$ expression. In a recent successful work on astronomical image restoration, Frieden[6] used the $-B\log B$ expression for his entropy.

Since these two schools have been working independently, the question arises as to which entropy expression is the correct one to use. This question was noted by Frieden[1] and Wernecke and D'Addario[2] as a challenging problem. The present paper is an effort to solve this problem by demonstrating the conditions under which the $\log B$ and the $-B\log B$ expressions should be used. We prove that these expressions are limiting cases of a more general expression. We are mainly concerned with the problem of image restoration (as worked on, for example, by Ponsonby,[5] Wernecke and D'Addario,[2] and Frieden[1,6]). However, our conclusions can be easily translated into the spectral power estimation problem (worked on, for example, by Burg,[3] Radoski, Fougere and Zawalick,[7] and Ulrych[8]).

The present paper is based on two basic interpretations of the ME restoration of an image. The first is that, when an image pattern is given, there are many different possible object patterns that can produce the same given image pattern. This many-to-one mapping from object to image is an intrinsic property of any real measuring system that has a finite aperture or bandwidth and hence a point-spread function of nonzero width. Another way of interpreting this point is the following. The information contained in the object pattern is carried by photons through an optical system and is recorded in the image. Information is lost in the optical system, and our problem is to analyze the nature of the optical communication channel, and then work out the restoration.

The first interpretation leads directly to the second. Since there are many possible object patterns, we are faced with the problem of choosing one or the problem of making a criterion for choosing one. The ME method, as interpreted in the present paper, provides such a criterion. It presupposes that a probability of occurrence can be assigned to each possible pattern, and then the ME method chooses the most probable pattern as the estimated, or restored, object pattern.

This second point of our interpretation of the ME restoration method may encounter resistance from some readers. Thus we hastily add some explanation. The word entropy is often interpreted either as the concept form which thermodynamic equilibrium is presupposed or as a measure of "randomness" (often without a strict definition of randomness). We do not invoke these concepts; in the present paper, the word entropy is used in the sense of Boltzmann's definition and is defined as the logarithm of the probability. Since the logarithm is a monotonic function of the argument, to find a maximum of the entropy is to find a maximum of the probability.

Based on these two fundamental concepts, we show in the present paper what the probability of finding an object pattern is, and then we show under what conditions the entropy form to be used is $\log B$ or $-B\log B$. Instead of starting from Shannon's entropy expression, as Burg[3] did, we go much further back. (Shannon's entropy is shown to be a special case in a later section.) We start (in Sec. II) with the probability of a brightness pattern in an object (based on the Bose-Finstein statistics, which photons obey). This leads to the $\log B$ and $-B\log B$ distinction for certain special cases. The number of degrees of freedom for the photons, a basic concept in

Sec. II, is discussed separately and in detail in Sec. III.

The entropy used in Sec. II is perfectly authentic and in keeping with Boltzmann's original concept; however, the formulas in Sec. II differ from the ordinarily seen $-f\log f$ form of the entropy. To resolve the questions some readers may have concerning these formulas a simple example of the multinomial expression applied to dice throwing is given in Sec. IV. The concepts of *a priori* probability and the familiar entropy form $-f\log f$ are explained in that section.

The dice example in Sec. IV leads to the next section, in which Shannon's and Burg's formalism are derived and interpreted. Section V presents the formulation based on the ensemble, which can be useful in accounting for noise in the ME formulation.

II. PROBABILITY OF AN OBJECT PATTERN

We assume that the object of interest is a two-dimensional distribution of photon sources in the far field. The two-dimensional space in which the object is viewed is divided by a hypothetical rectangular mesh into equal square cells of area w^2 each. The choice of the size w is arbitrary at this stage, but not without important consequences, as will be discussed below.

The number of photons coming from the jth object cell ($j=1,2,\ldots$) to the receiver in the observation time t is written as the dimensionless quantity n_j. (These photons may be emitted from the object or reflected from it.) The ultimate goal of the maximum entropy method we present in this paper is to calculate the most probable spatial pattern $\{n_j\}$ for the object. To free the presentation from unessential complications, we assume that the photons are quasimonochromatic with bandwidth $\Delta\nu$. Further, it is assumed that the object is not colorful, which implies that all $\Delta\nu_j=\Delta\nu$. This simplifies the discussions of coherence volume[9] and of the number of degrees of freedom for the photons.

A basic postulate of our analysis is that for each cell there correspond z degrees of freedom for the photons detected from that cell. The value z is proportional to the area w^2, the bandwidth $\Delta\nu$, the aperture of the detection apparatus, and the time interval t of observation. However, we first leave the value of z unspecified, except that it is a given positive integer and is constant. The meaning and the value of z are discussed in detail in the next section.

The number n_j of photons are arranged (distributed) over the z degrees of freedom within the bandwidth $\Delta\nu$ with the condition that multiple occupancy in one degree of freedom is allowed because of the Bose nature of photons. The number of macroscopically indistinguishable ways q_j that such an arrangement can be formed is expressed by the combinatorial formula for Bose-Einstein statistics[10]:

$$q_j(n_j)=\frac{(n_j+z-1)!}{n_j!\,(z-1)!}\quad . \tag{2.1}$$

Each of these arrangements is a quantum-mechanical state of the n_j photons. We now postulate that each of the q_j *different* (microscopically distinguishable but macroscopically indistinguishable) arrangements occurs with the same *a priori* probability. This postulate, which corresponds to the "equal-weight" principle of quantum mechanics that each eigenstate of the Hamiltonian is occupied with the same probability, is the basis of the statistical analysis of this paper. The equal-weight postulate allows interpreting q_j as the weight factor, or the degeneracy, for cell intensity n_j. Thus q_j is proportional to the probability that n_j photons are distributed over the z degrees of freedom.

The concept of the probability q_j in Eq. (2.1) can be understood by a simple example of throwing dice. Suppose we throw two dice and count the sum of the two numbers. The number of different ways the sum 5, for example, appears is four, because the possible combinations are (1,4), (2,3), (3,2), and (4,1). We write this as $q(5)=4$. The general expression for the number of ways $q(n)$ that the sum n appears is

$$q(n)=6-|7-n|\quad . \tag{2.2}$$

For fair dice, the probability that the sum n appears is proportional to $q(n)$. Corresponding to what we said concerning Eq. (2.1), the postulate for this example is that each of the four states (1,4), (2,3), (3,2), and (4,1) appears with the same *a priori* probability.

Different from $q(n)$ for dice in Eq. (2.2), the algebraic expression for $q_j(n_j)$ for photons in Eq. (2.1) contains factorials. For the purely practical reason that in working mathematically with factorials, it is easier to first take the logarithm, we take the logarithm of q_j in Eq. (2.1) and introduce an expression s_j:

$$s_j(n_j)\equiv\ln q_j(n_j)\quad . \tag{2.3}$$

We do not yet call this quantity entropy because we wish to avoid invoking the connotations and misconceptions sometimes associated with the word. We now examine some limiting cases of Eqs. (2.1) and (2.3).

[A] When $z=1$, Eqs. (2.1) and (2.3) give

$$q_j=1 \text{ and } s_j=0\quad . \tag{2.4}$$

This is understandable because, when the number of degrees of freedom is unity, all n_j photons occupy the same degree of freedom and, as photons are indistinguishable from one another, the degeneracy is unity.

[B] When $1<z\ll n_j$, it is easier to work with the logarithmic form $s_j(n_j)$ because we can use Stirling's approximation. Neglecting z/n_j in the expansion, we arrive at

$$s_j=(z-1)\ln n_j-\ln(z-1)!\quad . \tag{2.5}$$

[C] When $z\gg n_j$, we again use Stirling's approximation, neglect n_j/z, and approximate Eq. (2.3) as

$$s_j=n_j\ln z-n_j(\ln n_j-1)\quad . \tag{2.6}$$

This expression is exactly the classical (i.e., Maxwell-Boltzmann particle statistics) limit of Eq. (2.1) when n_j "particles" are distinguishable, in agreement with the general property that in the limit of $z\gg n_j$, Bose statistics approach classical statistics.

We are interested in the entire object made of many

cells, each of area w^2. We ask the number of ways n_1, n_2, ... photons come from the first, second, ... cells, respectively, independently of each other. Then the entire number of ways Q is the product of q_j's in Eq. (2.1) since each cell is independent:

$$Q(n_1, n_2, \ldots) = \prod_j q_j(n_j) \ . \tag{2.7}$$

We then see what the logarithm of Q looks like. In writing it, we introduce n for the total number of photons:

$$n = \sum_j^{\text{all } j} n_j \ , \tag{2.8}$$

and define, for mathematical convenience, a normalized object pattern

$$p_j \equiv n_j/n \ . \tag{2.9}$$

This introduction of the normalized p_j does not mean that we are assuming p_j to be independent of the total intensity of light n; when the object is illuminated by another light source, for example, p_j can vary as the total illumination changes because reflectivity may not be linear in intensity and uniform over the object. At any rate, it is meaningful to talk about the relative local intensity p for a given illumination condition, and for a given recorded image.

The three limiting cases of $S \equiv \ln Q$ are given below:

[A] When $z = 1$,

$$Q = 1 \text{ and } S = \ln Q = 0 \ . \tag{2.10}$$

This means that the probability Q is independent of the photon distribution $\{p_j\}$, whatever $\{p_i\}$ may be, and hence all distributions $\{p_j\}$ in the object space are equally likely!

[B] When $1 < z \ll n_j$, Eq. (2.5) leads to

$$S = \sum_j s_j = \sum_j [(z-1)\ln n - \ln(z-1)!\,] + (z-1)\sum_j \ln p_j \ . \tag{2.11}$$

[C] When $z \gg n_j$, Eq. (2.6) leads to

$$S = \sum_j s_j = n \ln(z/n) - n \sum_j p_j(\ln p_j - 1) \ . \tag{2.12}$$

Following Boltzmann[11] and Planck,[12] it is customary to call the logarithm of the degeneracy (weight factor, or probability) the entropy. Since the probability is proportional to the number of ways a certain event occurs (when each way appears with the same *a priori* probability), our S, which is the logarithm of the number of ways, may also be called the entropy.

In the "maximum entropy" algorithm for calculating the object distribution, z and the total intensity n are fixed numbers; therefore only the p-dependent terms in Eqs. (2.11) and (2.12) are significant. When we examine the p-dependent terms, we note that Eqs. (2.11) an and (2.12) contain $\sum \ln p_j$ and $-\sum p_j \ln p_j$, respectively. The p_j in these expressions is the normalized brightness which we wrote as B in the introduction. Therefore it is legitimate to identify case [B] with the $\log B$ expression used in ME image-restoration studies[2,5] and case [C] with the $-B\log B$ expression.[6] (We deliberately use B, for brightness, rather than p because of the opinion, sometimes heard, that the entropy is always of the form $-p\ln p$ and never $\ln p$.) In each of these expressions, an accurate value of z is not needed in finding the distribution $\{p_j\}$ that maximizes the entropy. What is needed is the knowledge of the ratio n/z. The number of photons n/z per degree of freedom (or per mode) is a useful concept and is widely used in the statistics of radiation.[13]

One exception to which the n/z criterion so far cannot apply is case [A], in which $z = 1$. The analysis in this section says that case [A], to which the wave analysis[3,7,8] initiated by Burg actually belongs, cannot be classified either to the $\log B$ case or the $-B\log B$ case. We return to this problem in Sec. V, in which we treat the problem based on a fixed value of the average of n rather than n itself.

III. NUMBER OF DEGREES OF FREEDOM

The idea of the number of degrees of freedom z for photons used in the previous section is basic in this paper. Therefore we discuss the concept in detail in this section. The z for photons emitted from an area w^2 of the object space and observed within time t can be developed in either real space or phase space.

In real space, first we consider the coherence volume for a photon. In the longitudinal direction of propagation, this volume may be considered to have a coherence length l given by the coherence time $\tau \sim 1/\Delta\nu$ multiplied by the velocity of propagation c:

$$l = c\tau \ . \tag{3.1}$$

In the transverse direction, the coherent area σ grows by propagation to a large range R and is inversely proportional to the area of the source w^2. Using an arbitrary cutoff for the acceptable degree of partial coherence, we may write a simple expression for σ from the Zernicke-van Cittert[14] theorem in the far-field limit:

$$\sigma \cong R^2\lambda^2/w^2 \ , \tag{3.2}$$

where λ is the wavelength of the photons. Again, note the dependence of σ on the source, except for R^2.

The coherence volume may be taken as the product of the coherence area σ and coherence length l:

$$v_{coh} = c\tau\sigma = c^3 R^2/\nu^2 \Delta\nu w^2 \ . \tag{3.3}$$

The coherence volume corresponds to one degree of freedom for the photon; it is not possible to distinguish the photons by interference experiments, for example, in one degree of freedom or in one coherence volume.

Now we define z. Suppose we detect photons in a detection time t. Ignoring the transverse coherence area for a moment, we can say that we have detected these photons in $t/\tau = z$ coherence lengths or degrees of freedom. The number of degrees of freedom should also depend upon the ratio of the area A of the detector aperture to the coherence area σ of a photon. This ratio contributes a factor A/σ in z, so z may be viewed as the number of times the coherence volume $c\tau\sigma$ (which represents a degree of freedom) is contained in the "detection volume" ctA:

$$z = \frac{ctA}{c\tau\sigma} = \left(\frac{t}{\tau}\right)\left(\frac{A}{\sigma}\right) \equiv z_{temp}\, z_{spat} \ . \qquad (3.4)$$

We may think of the number of degrees of freedom as being made of two factors: a temporal one z_t and a spatial (two-dimensional area) z_s. Whenever either z_t or z_s is less than unity, the photons detected are still indistinguishable, so we round z_t or z_s up to one degree of freedom.[9] We have used a simplified definition of the number of degrees of freedom to avoid nonessential complications; we assume that simple reciprocal relations hold between w^2 and σ, and between $\Delta\nu$ and τ, and that all $\Delta\nu_i = \Delta\nu$. A detailed study of coherence volume with more general partially coherent sources given in Ref. 9 agrees with the qualitative presentation given here.

The number of degrees of freedom can also be represented in phase space. The phase space of the conjugate variables, time, and frequency (bandwidth), is commonly employed in communication and signal processing theory Utilizing the sampling theorem, this phase space is usually partitioned into equal unit area slices at equal time intervals representing one degree of freedom ($z_t = 1$) per measurement in the sampling of a time series. Alternately, it is partitioned into equal unit area slices at equal frequency differences representing one degree of freedom ($z_t = 1$) per measurement in the sampling of a spectrum.

Since it is not necessary to have these particular partitionings or to have only one degree of freedom per measurement (except for the satisfaction of the sampling theorem), measurements are not always made in that fashion. As we shall see, for example, in the observation of Poisson temporal processes, z_t is necessarily greater than one.

The two variables, spatial extent, and spatial frequency which are conjugate to each other, define additional dimensions of phase space besides the familiar time-bandwidth dimensions. This geometrical phase space of space and reciprocal space has been used in studies[15] to define the number of degrees of freedom in images. This is our z_s factor. We include in the image analysis and estimation problem the previously neglected (in image studies) z_t to count the entire number of degrees of freedom in an object. Our previous definitions leading to the number of degrees of freedom [Eq. (3.4)] are simply a casting of phase space ideas in terms of coherent volumes and detection volumes.

Several comments are in order. Changing the arbitrary size w of the object space cells will have important effects on the reconstruction and possibly even on the statistics. This is not a defect of the method, but rather one of its important assets. For example, making w smaller than the resolution limit set by the detection aperture A will result in "super-resolution." When we make the object cell size w^2 smaller than a resolution element to achieve super-resolution, z_s remains at its minimum value of one degree of freedom,[9] but the number of photons decreases in proportion to w^2. This is an important example of a case where the n/z ratio is not a constant determined by physical properties of the source. The number of photons per degree of freedom decreases as the super-resolution increases as $A/\sigma < 1$ (and in the same fashion as $t/\tau < 1$). Super-resolution may be viewed as a special case of reconstructing from under-sampled data. More general cases of undersampling of the aperture plane (e.g., sparse arrays of antennas), which do not involve super-resolution at all in the reconstruction, can also be handled.

Throughout this discussion we have assumed rather ideal detection processes in which the photon distribution is closely mirrored by the distribution of photoelectrons, or of exposed silver halide grains in photography. The accumulation of photons has been assumed to be proportional to time. Similarly, we have assumed that the number of collected photons is proportional to the detector area A. When the efficiency of photoelectron detection is less than unity, binomial distributions for the probability of detection and nondetection alter the distributions.[16] However, for the estimation of n and n/z, it is sufficiently accurate to take the photoelectron distribution as a mirror of the photon distribution.[17] In fact, the distribution devised by Mandel[18] for photoelectrons from rather different considerations is exactly the negative binomial distribution which we also display for bosons. Photographic detection is complicated by a threshold effect, but the effects of photon statistics have been theoretically shown[19] to display the characteristic boson "clumping" effects with increasing n/z in the density versus exposure photographic characteristic curve, consistent with the Bose distribution. The distortions of detection will rarely be so great as to alter the estimate of the magnitude of n/z, which is all that is required here.

An absorption at the source or at the detector as well as change in the illumination intensity (in the case of reflection) not only change the total intensity, but may also influence the statistics. If we have *a priori* knowledge of this, it could be taken into account to determine a corrected n/z appropriate to the statistics before the lossy process.

Let us examine some examples which have different n/z values. If we had some *a priori* knowledge of the photon frequency distribution function of the object, for example, if we know that the radiation detected at the frequency ν was from a black body at a certain temperature T, then we could use the well-known relation for the average number of photons per mode (or per degree of freedom) at thermal equilibrium n/z as a function of T and ν:

$$n/z = [\exp(h\nu/kT) - 1]^{-1} \ , \qquad (3.6)$$

where h is Planck's constant. A reduced value of n/z is employed when $A/\sigma < 1$, as in super-resolution, or when $t/\tau < 1$ as mentioned before. As an illustration, suppose we were to restore an image of the sun, which is approximately a blackbody, the mean chromosphere temperature of which could be taken closely at 6000 °K. Using Eq. (3.6), we find that $n/z \ll 1$ in the visible region and that $-B\log B$ is the appropriate limiting form of the entropy expression. For a solar image using wavelengths larger than 3 μm, n/z is greater than one, and the correct form would be $\log B$. Since all the planets have a very small n/z ratio in the visible and infrared

regions, $-B\log B$ is the appropriate entropy form to use for restoring their images.

Without this kind of blackbody *a priori* information, we could still make independent measurements of z or of the n/z ratio. A determination of source size, $\Delta\nu$ from the spectral line shape, and detector parameters will give z. The number of photons n is presumed measurable. More conveniently, if the isotropic flux or the related brightness B (W m^{-2} sterad^{-1} Hz^{-1})is measured, then

$$n/z = B(\nu)\, c^2/2h\nu^3 \ . \tag{3.7}$$

The factor of 2 is chosen for the case of unpolarized brightness. This relation, of which the blackbody distribution is a special case, can be seen to follow from the definition of B and from Eqs. (3.2), (3.3), and (3.4), providing $1/\tau$ is simply $\Delta\nu$. To this approximation we may apply Eq. (3.7) to estimate n/z for nonthermal sources such as the many radio astronomical objects where synchrotron radiation, for example, plays a dominant role. Equation (3.7) tells us that the bright Double Cygnus A (3C 405) at 960 MHz has an enormous $n/z \sim 10^6$. Similarly, Centaurus A (CTA 59) at 178 MHz has a very large $n/z \sim 10^4$. The bright radio astronomical objects likewise have a large n/z, and the log$B$ entropy expression is appropriate.[20] Images of less bright objects and at shorter wavelengths will dictate the $-B\log B$ form.

For the optical astronomical objects Frieden worked with, we see $n/z \ll 1$ and hence $-B\log B$ is the entropy expression that Eq. (2.12) gives, which is in agreement with Frieden's choice. The radio astronomical case of Wernecke satisfies the relation $n \gg z > 1$, and hence the log$B$ expression of Eq. (2.11) is the right entropy expression, which again supports Wernecke's choice.

IV. EXAMPLE OF MANY THROWS OF DICE

The main purpose of the present paper is to point out the importance of the $q_i(n_i)$ factor in Eq. (2.1) in formulating the entropy of the ME method. However, we must still answer the legitimate question of why our initial and fundamental entropy (2.1) does not resemble the expression ordinarily seen for entropy, namely $-\sum f_n \ln f_n$ or $-\sum p_i \ln p_i$ where, for example, f_n is the distribution function and p_i might be the probability. To answer this question, we start with an example of counting the numbers on many throws of two dice.

In the existing log$B$ school of ME spectral estimation, for example, the basic assumption is that the measured number of photons n can be taken as the estimate of the *average* number of photons $\bar{n}_i$ in unit spectral width (or emitted from the ith cell of the object). This is different from the basic postulate of Sec. II that the number (not the average number) of photons of the ith cell is n_i. When we say the average is $\bar{n}_i$, we are thinking of many cases with varying individual numbers. For example, when two dice are thrown many times, the average of the throws is 7, but individual throws can be distributed between 2 and 12. To explain the relation between the point of view of Sec. II, which is based on one fixed number n_i, and the other point of view based on the average $\bar{n}_i$, we use yet another dice example in this section.

Suppose we throw two dice many times and ask how often the number n appears. Let the total number of throws be M, and the number of throws in which n appear be Mf_n. Then the probability that, out of the entire M throws, 2 appears Mf_2 times, 3 appears Mf_3 times, . . . , is written

$$\mathcal{P}(f_2, f_3, \ldots, f_{12}) = \frac{M!}{(Mf_2)!\,(Mf_3)!\ldots(Mf_{12})!} \times q(2)^{Mf_2} q(3)^{Mf_3} \ldots q(12)^{Mf_{12}}$$

$$\sum_{n=2}^{12} f_n = 1 \ , \tag{4.1}$$

in which $q(n)$ is the probability that n is found in one throw. The previous equation (2.2) is an example of $q(n)$ for a pair of fair dice.

When we throw the dice a very large number of times ($M \gg 1$), the number of times n appears approaches $Mf_n^{(o)}$, where the set $\{f_n^{(o)}\}$ maximizes the probability function $\mathcal{P}(f_2^{(o)}, f_3^{(o)}, \ldots, f_{12}^{(o)})$. The common mathematical procedure of finding a maximum of such a function is to require derivatives of its logarithm to vanish. So we first form the logarithm using Stirling's approximation (which is permitted since M is a large number) as

$$\mathcal{S} \equiv \ln \mathcal{P} = M\left[-\sum_{n=2}^{12} f_n \ln f_n + \sum_{n=2}^{12} f_n \ln q(n) + \lambda\left(1 - \sum_{n=2}^{12} f_n\right)\right] . \tag{4.2}$$

The last terms, $\lambda(\ )$, are written to satisfy the normalization of f_n, and λ is the Lagrange multiplier. When we maximize $\mathcal{S}$ in Eq. (4.2), we arrive at the most probable distribution:

$$f_n^{(o)} = q(n) \Big/ \sum_{n=2}^{12} q(n) \ . \tag{4.3}$$

Several comments are in order.

(i) The $\mathcal{P}(f_2, f_3, \ldots, f_{12})$ expression in (4.1) is a term in the multinomial expression. The expansion can be written

$$[q(2) + q(3) + \ldots + q(12)]^M = \sum \mathcal{P}(f_2, f_3, \ldots, f_{12}) \ , \tag{4.4a}$$

where the sum $\sum$ is done under the condition

$$Mf_2 + Mf_3 + \ldots + Mf_{12} = M \ . \tag{4.4b}$$

The multinomial expression $\mathcal{P}(f_2, f_3, \ldots, f_{12})$ is used extensively in probability theory.[21]

(ii) The probability expression $\mathcal{P}(f_2, f_3, \ldots, f_{12})$ in Eq. (4.1) is based on M throws of the dice. A collection of repetitions of an event is often called an ensemble. Thus we can say f_n represents the distribution of n over an ensemble.

As we see in Eq. (4.3), the probability of a single throw $q(n)$ is proportional to the distribution function over an ensemble. In the same way, any probability

consideration (for example, the probability concept used in Sec. II) has the concept of an ensemble in its foundation. Actually the probability can *only* be defined using many tries [this is discussed in (iii) below]. However, when we use the term ensemble in the present paper, it is defined in a limited sense, namely that the average number of photons $\bar{n}$ is fixed in the ensemble, but that individual n's in the individual facsimiles can fluctuate.

(iii) For fair dice, we can write on $q(n)$ of Eq. (4.3) as given in Eq. (2.2). However, if the dice are loaded, the probability of finding n is not given by Eq. (2.2), but can be a complicated function of n. Even in such a case, many throws (the limit of $M \to \infty$) can yield the knowledge of $q(n)$ through Eq. (4.3). In Sec. II, we used the quantum nature of photons and accepted the formula for $q_j(n_j)$ as given in Eq. (2.1).

(iv) The reason why we give the dice example in this section is to point out that the probability expression $\mathcal{P}(f_2, \ldots, f_{12})$ in (4.1) is made of two factors. We will call the combinatorial factor $\mathcal{P}_1$ and the rest $\mathcal{P}_2$. It is a common practice, which we do not follow in this paper, to call $\ln\mathcal{P}_1$ the entropy; $\ln\mathcal{P}_1$ takes the familiar form

$$\ln\mathcal{P}_1 = -M\sum_n f_n \ln f_n \ . \tag{4.5}$$

The most probable distribution $f_n^{(0)}$ is derived *not* by maximizing only $\ln\mathcal{P}_1$ in Eq. (4.1) but by maximizing the entire probability $\ln\mathcal{P}_1 + \ln\mathcal{P}_2$ in Eq. (4.2). Only in the case in which $\mathcal{P}_2$ is a constant, independent of f_n, can the most probable $f_n^{(0)}$ be derived by maximizing $\ln\mathcal{P}_1$ alone. This latter case will be discussed as a special case of the development given in the next section.

V. FORMULATION BASED ON THE AVERAGE $\bar{n}_j$

In Sec. II we mentioned under the classification [A] that when $z = 1$ we cannot reconstruct the image using the ME concept (and Burg's case belongs to this class). This is because the probability of n_j photons being emitted from the jth cell is calculated using the $q_j(n_j)$ expression of (2.1), which is identically unity for one degree of freedom.

Even when $z = 1$, however, we can talk about the *probability* that n photons come from the cell if we have previous knowledge, by some means, of the distribution function f_n. Actually we do not need to know the entire function f_n, but it is sufficient to identify n_j as an estimate of the average number $\bar{n}_j$ of the distribution. When we make this identification, we can derive the most probable distribution $f_n^{(0)}$ and the probability of finding $\bar{n}$ associated with $f_n^{(0)}$. Thus, we formulate the present section based on the distribution and on the concept of the ensemble (in the sense we defined it in the previous section). This treatment allows us to understand some cases of the log$B$ school of the ME estimation method.

We consider an ensemble made of a large number M of nearly identical facsimiles or sytems, each system representing the single jth cell treated in Sec. II. The number of systems which have n photons each is written $f_n M$. This function f_n is not the so-called cumulative distribution, and it is sometimes called the probability density function. This definition implies the normalization for f_n:

$$\sum_{n=0}^{\infty} f_n = 1 \ . \tag{5.1}$$

The set $\{f_n\}$ specifies a particular state of the ensemble.

We are interested in counting the number of *different* ways, which we designate as $P\{f_n\}$, that an ensemble can be constructed. $P\{f_n\}$ is made of two factors. The first is the number of ways of making a distribution $\{f_n\}$ over the ensemble:

$$\Omega\{f_n\} = M! \Big/ \prod_n (Mf_n)! \ . \tag{5.2}$$

This corresponds to the combinatorial factor in Eq. (4.1). The second factor comes from the fact that n photons from a cell have the *a priori* weight or degeneracy factor $q(n)$ as is represented in Eq. (2.1):

$$q(n) = (n + z - 1)!/n!(z - 1)! \ . \tag{5.3}$$

Since each system can take its configuration independently from the other systems in the ensemble, the second factor $Q\{f_n\}$ is a product of (5.3) for each system and is expressed

$$Q\{f_n\} = \prod_n q(n)^{Mf_n} \ . \tag{5.4}$$

The total number of ways that the distribution $\{f_n\}$ is achieved in the ensemble is then the product of the two factors $\Omega\{f_n\}$ in (5.2) and $Q\{f_n\}$ in (5.4):

$$P\{f_n\} = \Omega\{f_n\} Q\{f_n\} \ . \tag{5.5a}$$

This P is of exactly the same nature as $\mathcal{P}$ in Sec. IV and is a multinomial expression:

$$P\{f_n\} = \frac{M!}{\prod_n (Mf_n)!} \prod_n q(n)^{Mf_n} \ . \tag{5.5b}$$

After we thus define the distribution $\{f_n\}$ over the ensemble, we ask for the most probable distribution $\{f_n^{(0)}\}$ when the average n is given as $\bar{n}$. This distribution $\{f_n^{(0)}\}$ must be the one which maximizes the probability (degeneracy, or the number of ways of constructing the ensemble) $P\{f_n\}$ in (5.5). Since it is convenient for mathematical reasons, we maximize $\ln P\{f_n\}$:

$$S\{f_n\} \equiv \ln P\{f_n\} = \ln\Omega\{f_n\} + \ln Q\{f_n\} \ . \tag{5.6}$$

One may call either $\ln\Omega\{f_n\}$ the entropy or the entire $\ln P\{f_n\}$ the entropy. However, there is no such arbitrariness when the maximum distribution $\{f_n^{(0)}\}$ is to be derived. The function to be maximized is not $\ln\Omega\{f_n\}$, which has the familiar $-f\ln f$ form of entropy, but the entire $\ln P\{f_n\}$, which includes the *a priori* terms. Since the term "maximum *entropy* restoration" is in use, we will call the entire expression (5.6), $\ln P\{f_n\}$, the entropy in this paper.

Using Stirling's approximation, we derive

$$S\{f_n\}/M = -\sum_n f_n(\ln f_n - 1) - 1 + \sum_n f_n \ln q(n) + \mu\left(\bar{n} - \sum_n n f_n\right) + \lambda\left(1 - \sum_n f_n\right) \ . \tag{5.7}$$

In anticipation of the maximization, we used Lagrange multipliers λ and μ for the subsidiary conditions expressed by Eq. (5.1) and by

$$\bar{n} = \sum_n n f_n \ . \tag{5.8}$$

The most probable distribution $f_n^{(o)}$ is derived by maximizing (5.7) as

$$f_n^{(o)} = e^{-\lambda - \mu n} \frac{(n+z-1)!}{n!\,(z-1)!} \ . \tag{5.9}$$

The normalization coefficient $e^{-\lambda}$ is determined from Eq. (5.1) as

$$e^{\lambda} = \sum_{n=0}^{\infty} \frac{(n+z-1)!}{n!\,(z-1)!} e^{-\mu n} = \frac{1}{(1-e^{-\mu})^z} \ . \tag{5.10}$$

Because the combinatorial factor in (5.9) is thus of the expansion coefficients in (5.10), the distribution in (5.9) is called the negative binomial distribution,[22] and (5.10) is its generating function. The other multiplier μ is determined from (5.8) as

$$\mu = \ln \frac{\bar{n}+z}{\bar{n}} \ . \tag{5.11a}$$

Substituting this in (5.10) allows us to write λ as

$$\lambda = z \ln \frac{\bar{n}+z}{z} \ . \tag{5.11b}$$

Using these two expressions in (5.9), we can write $f_n^{(o)}$ explicitly as

$$f_n^{(o)} = \left(\frac{z}{\bar{n}+z}\right)^z \left(\frac{\bar{n}}{\bar{n}+z}\right)^n \frac{(n+z-1)!}{n!\,(z-1)!} \ . \tag{5.12}$$

When the entropy expression (5.7) is a maximum and f_n is equal to $f_n^{(o)}$ in (5.9), we can show, using Stirling's approximation, that the entropy corresponding to the most probable distribution is

$$S\{f_n^{(o)}\}/M = \mu\bar{n} + \lambda = (\bar{n}+z)\ln(\bar{n}+z) - \bar{n}\ln\bar{n} - z\ln z \ . \tag{5.13}$$

An entropy expression equivalent to (5.13) was derived by Gamo[23] for partially coherent light beams.

We examine the three limiting cases corresponding to those presented in Sec. II.

[A] When $z = 1$, the *a priori* probability $q(n)$ in (5.3) reduces to unity so that the general entropy expression (5.7) simplifies to the first summation $-\sum f_n \ln f_n$; this is the expression of the entropy used by Shannon in his information theory. In this case, the most probable distribution (5.12) is reduced to the exponential distribution[24]:

$$f_n^{(o)} = \frac{1}{\bar{n}+1}\left(\frac{\bar{n}}{\bar{n}+1}\right)^n \ , \tag{5.14}$$

and the entropy expression (5.13) becomes

$$S\{f_n^{(o)}\}/M = (\bar{n}+1)\ln(\bar{n}+1) - \bar{n}\ln\bar{n} \ . \tag{5.15}$$

When $\bar{n} \gg 1$, this further reduces to

$$S\{f_n^{(o)}\}/M = 1 + \ln\bar{n} \ . \tag{5.16}$$

This equation is the classical wave (field) limit of the entropy expression for the Bose-Einstein statistics.

The entropy expression in (5.16) is equivalent to the form that Burg used in his ME treatment. (The mathematics is the same if we formulate for the spectral analysis problem.) Our derivation of (5.16) in this section is similar to the method used in Shannon's theory[4] of communication. In Shannon's treatment, the frequency of repetition of measurement is controlled by the sampling theorem; each measured sampled point therefore corresponds to one degree of freedom, mode or state, and hence $z = 1$. Naturally, the Gaussian distribution in complex amplitudes for the classical wave representation or, equivalently, for the exponential distribution of photons plays the dominant role in the traditional information theoretical exposition as the most probable distribution, and for these cases our reasoning supports Shannon's expression of information.

In considering the distribution function f_n in this treatment, Shannon considers n to be a continuum variable. We can easily check that, when summations over discrete n used in deriving (5.15) are changed into integrations over the continuum of n, we arrive at (5.16) rather than (5.15). So far as our formulation is concerned, this is equivalent to assuming $\bar{n} \gg 1$.

In deriving the negative exponential distribution (5.14), we assumed that $\bar{n}$ is fixed. This procedure is equivalent to using the complex wave amplitude formulation together with the assumption that the power (i.e., the second moment) is given and deriving the Gaussian distribution in complex amplitude as was done by many authors in communication, astronomical imaging, and geophysical applications.[25]

It is noteworthy that the distribution function (5.12) for a general z (when more than one degree of freedom is involved in the considerations) can be derived from the $z = 1$ case (5.14). As an example, we derive the case for $z = 2$. The distribution function for this case is derived using the property that the distributions in each degree of freedom are independent and thus the joint probability is a sum of products of (5.14) in the convolution form:

$$\sum_{n=0}^{N} f_n^{(o)} f_{N-n}^{(o)} = \left(\frac{\bar{N}}{\bar{N}+1}\right)^2 \left(\frac{\bar{N}}{\bar{N}+1}\right)^N (N+1) \ . \tag{5.17}$$

By comparing this with (5.12), we can verify that (5.17) is the distribution $f_n^{(o)}$ in (5.12) for the case $z = 2$ and $\bar{N} = 2\bar{n}$. Using (5.10) and the generation function technique, it can be found in general that, when z factors of the exponential distribution (5.14) are convoluted, the negative binomial distribution (5.12) results with the mean of the resultant distribution equal to z times the original average $\bar{n}$ in (5.14). This convolution is another way of understanding the negative binomial distribution (5.12) that arises when more than one degree of freedom is involved. This result also confirms the legitimacy of the entropy expression (5.6) in which both the Ω term and the Q term are included.

[B] When $z \gg \bar{n}$, the entropy expression (5.13), using Stirling's approximation, reduces to

$$S\{f_n^{(o)}\}/M = z\ln\bar{n} + z - z\ln z \ . \tag{5.18}$$

This corresponds to the $\log B\,(\ln\bar{n})$ entropy expression (2.5), recalling that z is a constant, and $\bar{n}$ is the vari-

able. The other terms are merely additive constants.

[C] When z is large compared with n, we can use the same approximation as we used in (2.6) and reduce $f_n^{(o)}$ in (5.9) to

$$f_n^{(o)} = e^{-\lambda-\mu n} z^n/n! \quad . \tag{5.19}$$

Further, (5.11) reduces to

$$\mu = \ln(z/\bar{n}) \text{ and } \lambda = \bar{n} \quad . \tag{5.20}$$

From these two, we obtain

$$f_n^{(o)} = e^{-\bar{n}}(\bar{n})^n/n! \quad , \tag{5.21}$$

which is the Poisson distribution. The entropy expression (5.13) reduces to

$$S\{f_n^{(o)}\}/M = \bar{n}\ln z - \bar{n}(\ln\bar{n} - 1) \quad . \tag{5.22}$$

This corresponds to the $-B\log B$ expression in (2.6), again recalling that z is a constant and $\bar{n}$ is the variable. Equation (5.22) is the classical (i.e., Maxwell-Boltzmann) particle limit of the Bose statistics.

The Poisson distribution of photons in this limit is well-established experimentally. The derivation of the Poisson distribution in this limit [C] and the exponential distribution in limit [A] supports the correctness of the analysis in this section. Many authors have noted the importance of the Poisson distribution and have employed it for imaging[6] and communication[26] studies. However, they did not derive it as a *most probable* distribution under specific conditions as they did for the Gaussian, but adopted the Poisson distribution *ad hoc*. The reason they[25] could not so derive the Poisson distribution is that they maximized only the $\ln\Omega\{f_n\}$ term, leaving out the $\ln Q\{f_n\}$ term in (5.6). In evaluating the probability when f_n is of the Poisson form, they again incorrectly used $-\sum f_n \log f_n$ for the (logarithm of) probability (which does not represent the total probability in the Poisson case). However, Rényi and McFadden[27] did each independently derive the Poisson distribution by maximizing the probability expression, which included the appropriate *a priori* term, as we do in the present paper. Although they were interested only in the properties of the probability of point processes, their approach and result support our reasoning.

One objection to including the $\ln Q\{f_n\}$ term in the entropy expression might be that such a definition would not satisfy the intuitively motivated induction condition, the third property for the measure of information postulated by Shannon: namely, that if an original choice is composed of several successive choices, then the measure of information should be expected to be the weighted sum of the individual measures.[28] On examination, this condition or axiom is meaningful only when the choices are, *a priori*, equally likely; it is not suitable when they are not. When the choices are equally likely, the $\ln Q\{f_n\}$ term in (5.6) is irrelevant and thus the $-f\log f$ form of Shannon's information results.

VI. RELATION BETWEEN THE TWO DERIVATIONS

In Secs. II and V, we derived the entropy expressions for a cell and for an ensemble of cells, respectively. By comparing the two results, we can show the consistency of our analysis. We will demonstrate that if the exposure time of forming the image is very long, for example, the analysis of Sec. II (which is based on a fixed number of photons) and the analysis of Sec. V (which is based on a fluctuating number of photons) agree with each other in the entropy expression as long as n/z is the same in both.

We rewrite (2.1) for one cell as

$$s_c = \ln \frac{(n_c + z_c - 1)!}{n_c!\,(z_c - 1)!} \quad , \tag{6.1}$$

where we write n_c and z_c for n_j and z, respectively, in (2.1), the subscript c indicates that these are quantities for a cell. The ensemble expression (5.13) is written as it is:

$$S = M[(\bar{n} + z)\ln(\bar{n} + z) - \bar{n}\ln\bar{n} - z\ln z] \quad . \tag{6.2}$$

In comparing the two expressions, n_c is the number of photons detected from the cell of the area a^2 within a certain specified detection volume $V = ctA$ [Eq. (3.4)]. The quantity z_c is the number of degrees of freedom associated with a photon coming from that cell. There is no restriction on the values of n_c and z_c, except that they be non-negative integers. It is meaningful to ask about the relationship between (6.1) and (6.2). For this purpose, let the detection volume V used in defining z_c in (6.1) be equal to M times the corresponding V used in (6.2). This introduces the correspondence

$$n_c = M\bar{n} \text{ and } z_c = Mz \quad . \tag{6.3}$$

Substituting (6.3) into (6.1) brings s_c exactly into the form (6.2) for S when we use the approximation

$$Mz \gg 1 \tag{6.4}$$

together with Stirling's approximation (which is justified when M is very large). It is to be noted particularly that the limit of $z_c \gg 1$ in (6.1) agrees with the case of $z = 1$ in (6.2).

VII. FERMI STATISTICS

The formulation in this paper has been based on photons and the Bose-Einstein statistics. It is natural to ask about the case of the Fermi-Dirac statistics because an image-restoration problem of a similar kind exists there (for example, in the field of electron microscopy).

For the Fermi case, we need modify Sec. II only slightly. We again divide the object space into cells. The number of electrons which have been emitted from the jth cell within the observation time t is written as n_j. The number of degrees of freedom corresponding to the cell is written as z. The number of ways q_j that the n_j electrons can be distributed over the z degrees of freedom is written by the binomial expansion coefficient (Fermi-Dirac statistics)

$$q_j = z!/(z - n_j)!\,n_j! \quad , \tag{7.1a}$$

which replaces Eq. (2.1). We can define the entropy for the cell as we did in (2.3):

$$s_j = \ln q_j \quad . \tag{7.1b}$$

As in Sec. II, case [A], $z = 1$, yields $s_j = 0$. Different

from previous sections, the restriction in the Fermi case is

$$z \geq n_j \ , \tag{7.2}$$

because of exclusion principle, and hence case [B] in Sec. II is of no concern. Corresponding to case [C], $z \gg n_j$, the entropy expression for the entire object field becomes

$$S = \sum_j s_j = n \ln(z/n) - n \sum_j p_j (\ln p_j - 1) \ , \tag{7.3}$$

where we define, as in (2.8),

$$n = \sum_j^{\infty} n_j \ , \tag{7.4a}$$

$$p_j = n_j/n \ . \tag{7.4b}$$

Since case [C] is the classical particle limit, it is natural that the entropy expression (7.3) is exactly the same as that derived for the classical particle limit of the Bose case (2.12). The ensemble formulation can be done as in Sec. V by replacing $q(n)$ in (5.3) with the binomial expression corresponding to (7.1a). Again for case [C], in which $z \gg \bar{n}$, we arrive at the classical particle expression (5.22).

VIII. DISCUSSION

In clarifying the concept of entropy discussed in previous sections, it is important to comment on the entropy as used in a class of problems that is similar to, but distinctly different from, ours. In statistical mechanics as well as in other fields, the entropy can always be written $\ln\Gamma$, where Γ is the number of microscopically distinguishable different ways in which the macroscopic configuration of the system can be arranged under given constraints defined in the problem, whatever the problem may be. This concept of entropy $\ln\Gamma$ is valid for a state that has fluctuated away from the most probable one as well as for the most probable one, and is a generalization of thermodynamic entropy that is defined only for the equilibrium state.

When the problems are different, the entropies are different. Sometimes confusions occur, however, when expressions of entropy which resemble each other are used in different problems. We will discuss an example that has caused confusion in understanding the maximum entropy image-restoration problem. In treating speckle patterns formed by lasers, Dainty[29] introduces the probability density function $p(I)$ of light intensity I, independent of where in the pattern I exists, and writes the entropy as

$$S = -\int p(I) \ln p(I) \, dI \ . \tag{8.1}$$

This is the legitimate entropy function for his problem when no contribution from the *a priori* probability is assumed. But the question he is trying to answer is different from ours and hence the physical content of the entropy expression (8.1) is different from ours.

In the speckle pattern statistics, in deriving (8.1) one calculates the number of ways Γ_D that all the different spatial patterns can be formed consistent with the unknown intensity probability distribution $\{p(I)\}$ to be determined and writes the entropy $\ln\Gamma_D$. This formulation of the problem does not lead to a unique two-dimensional pattern corresponding to a maximum of the entropy. Many patterns can be found, each of which satisfies the calculated distribution function corresponding to the maximum entropy.

In the problem of maximum entropy image restoration, on the other hand, the value of the intensity of each cell over a two-dimensional space is given an assumed fixed trial value. We ask in how many ways Γ_i the intensity at the cell i can be formed by taking into account the *a priori* degeneracy associated with the intensity at the point and any ensemble contribution if appropriate. The entropy is then derived as $\ln\Gamma_i$ for the cell i and then summed over all cell positions over the two-dimensional space. This is what we explained in Secs. II and V. The entropy is then maximized and a unique two-dimensional pattern is derived corresponding to it.

The basic difference between the speckle pattern statistics problem and the maximum entropy image problem is, therefore, that in the latter a unique pattern is calculated corresponding to the maximum entropy, whereas in the former a group of many patterns is derived from the intensity distribution corresponding to the maximum entropy of that problem. This difference leads to another important consequence with regard to the "smoothness" and the "disorder" of the pattern when the entropy is maximized. Suppose one asks for the pattern corresponding to the maximum of the entropy in the speckle statistics example. When (8.1) is maximized with only the constraint that the average $\langle I \rangle$ of I is given (together with the normalization constraint), then the result is that $p(I) = \langle I \rangle^{-1} \exp(-I/\langle I \rangle)$. This means the local intensity (I) can take various different values and hence the actual patterns are not "smooth." On the other hand, when the entropy of the maximum entropy image-restoration problem is maximized with only the constraint that the average intensity is given (p_i's are normalized), the local intensities p_i (i indicating the ith cell location) are constant and independent of i, resulting in a "smooth" flat pattern with the same intensity everywhere. Such a flat pattern in the speckle statistics would be expressed $p(I) = \delta(I - I_0)$ rather than the negative exponential distribution. This is only an illustration of the difference; it is not meant to imply that the one method always gives "smooth" and the other method "not smooth" results. Different constraints could reverse the situation. We avoid discussing the poorly defined concept of "disorder" in the context of two-dimensional patterns. Disorder should not be confused with the degree of smoothness, as they are independent concepts.

The two problems compared in this section use the same basic concept of the entropy $\ln\Gamma$, but the way the maximum entropy behaves seems quite different; this is only because the problems are different.

IX. SUMMARY AND CONCLUSION

In the maximum entropy (ME) image-restoration formulation, there are two different expressions for entropy now in use; for short, we call them the $\log B$ and the

$-B\log B$ expressions (B being the local brightness of the object, or the special spectral energy). In the present paper, we develop a general form for entropy that should be used in the ME restoration method, and classify the conditions under which the above two expressions are valid.

We use the original definition of entropy by Boltzmann and Planck, and find the most probable object pattern (restricted by the measured image). In other words, the *entropy* to be maximized is the logarithm of the *probability* that an object pattern occurs.

The object pattern is defined by the distribution of the number of photons n_i over the cells $i = 1, 2, \ldots$ in the object space. The probability $[q_i(n_i)]$ mentioned in the preceding paragraph is discussed in Sec. II, and is based on the Bose-Einstein statistics, which photons obey. It is proportional to the number of ways the n_i photons (in the ith object cell) can be distributed over z degrees of freedom:

$$q_i(n_i) = \frac{(n_i + z - 1)!}{n_i!\,(z-1)!} \quad .$$

Section III discusses in detail how z can be estimated. Based on this expression for $q_i(n_i)$, Sec. II interprets the two entropy expressions as the two limiting cases: [B] when $n_i \gg z > 1$, $\log q_i(n_i)$ leads to the $\log B$ expression for the entropy, and [C] when $z \gg n_i$, it leads to the $-B\log B$ form of entropy.

The relation between the entropy $s_i = \log q_i(n_i)$ and the entropy familiar in information theory is discussed in Sec. V, using the ensemble for which $\bar{n}_i$, an average over the ensemble, is specified rather than n_i. With $\bar{n}$ specified, the two distinctions [B] and [C] above remain the same. The third case, [A], when $\bar{n}_i \gg z = 1$ (the current applications of the ME method for power spectral estimation belong to this limit) leads to the $\log B$ expression when the ensemble formulation is used.

[1] B. R. Frieden, in *Picture Processing and Digital Filtering*, edited by T. S. Huang (Springer-Verlag, New York, 1975).

[2] S. J. Wernecke and L. R. D'Addario, "Maximum Entropy Image Reconstruction," IEEE Trans. Computers **C-26**, 351 (1977); S. J. Wernecke, "Two-Dimensional Maximum Entropy Reconstruction of Radio Brightness," Radio Sci. (to be published).

[3] J. P. Burg, project scientist, "Analytical Studies of Techniques for the Computation of High-Resolution Wavenumber Spectra," prepared by T. E. Barnard, Texas Instruments, Advanced Array Research Special Report No. 9, Contract No. F33657-68-C-0867, May 14, 1969.

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois, Urbana, 1949).

[5] J. E. B. Ponsonby, "An Entropy Measure for Partially Polarized Radiation and its Application to Estimating Radio Sky Polarization Distributions from Incomplete 'Aperture Synthesis' Data by the Maximum Entropy Method," Mon. Not. Roy. Astron. Soc. **163**, 369–380 (1973).

[6] B. R. Frieden and D. C. Wells, "Restoring with Maximum Entropy III: Poisson Sources and Backgrounds," J. Opt. Soc. Am. (to be published).

[7] H. R. Radoski, P. F. Fougere and E. J. Zawalick, "A Comparison of Power Spectral Estimates and Applications of the Maximum Entropy Method," J. Geophys. Res. **80**, 619–625 (1975).

[8] T. Ulrych, "Maximum Entropy Power Spectrum of Long Period Geomagnetic Reversals," Nature **235**, 218–219 (1972).

[9] A. Zardecki, C. Delisle and J. Bures, *Coherence and Quantum Optics*, edited by L. Mandel and E. Wolf, Proc. 3rd Rochester Conf., June 1972 (Plenum, New York, 1973).

[10] R. W. Ditchburn, *Light*, 3rd ed. (Interscience, New York, 1976), p. 697, or G. R. Fowles, *Introduction to Modern Optics* (Holt Rinehart & Winston, New York, 1968), p. 211.

[11] Ludwig Boltzmann, *Vorlesunger über Gastheorie* (J. A. Barth, Leipzig–Part I, 1896; Part II, 1898), translated by Stephen G. Brush as *Lectures in Gas Theory* (University of California, Berkeley, 1969). See pp. 5off., 74ff. and 371, where Boltzmann refers to the proportionality between entropy and the logarithm of the probability of a state.

[12] M. Planck, *Theory of Heat Radiation*, translated by M. Masius (Blakiston's, Philadelphia, 1914), Part III, Chap. I. The equation $S = k\log W$, which is now referred to as Boltzmann's Principle, first appears in this work.

[13] See, for example, L. Mandel, in *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1963), Vol. II, L. Mandel and E. Wolf, "Coherence Properties of Optical Fields," Rev. Mod. Phys. **37**, 231–287 (1965). The quantity $n.z$ is often referred to in these references and elsewhere in the literature as the degeneracy parameter of the radiation. We deliberately avoid this usage to prevent a later semantic confusion.

[14] J. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968), p. 138.

[15] See, for example, E. L. O'Neill and T. Asakura, "Optical Image Formation in Terms of Entropy Transformations," J. Phys. Soc. Japan **16**, 301–308 (1961).

[16] M. Scully, in *Quantum Optics Course XLII*, edited by R. J. Glauber (Academic, New York, 1969), p. 620ff. A succinct statement of the problem in terms of the coherent state representation. See also Ref. 14.

[17] J. Peřina, in *Quantum Optics*, edited by S. M. Kay and A. Maitland (Academic, New York, 1970).

[18] L. Mandel, "Fluctuations of Photon Beams: the Distribution of the Photo-Electrons," Proc. Phys. Soc. **75**, 233–243 (1959).

[19] W. M. Rosenblum, "Effect of Photon Distributions on Photographic Grain," J. Opt. Soc. Am. **58**, 60–62 (1968).

[20] Calculated from data given in J. Kraus, *Radio Astronomy* (McGraw-Hill, New York, 1966).

[21] R. von Mises, *Mathematical Theory of Probability and Statistics* (Academic, New York, 1964), Chap. IV, Sec. 3.2.

[22] The negative binomial distribution was first introduced by F. Eggenberger and G. Pólya, Zeits. Ang. Math. Mech. **3**, 279 (1923). According to Mandel, it was first applied to this problem by R. Fürth, Z. Phys. **48**, 323 (1928); **50**, 310 (1928).

[23] H. Gamo, "Thermodynamic Entropy of Partially Coherent Light Beams," J. Phys. Soc. Japan **19**, 1955–1961 (1964).

[24] The exponential distribution is sometimes referred to as the geometrical or the "pure Bose" distribution in the literature.

[25] See, for example, the extensive bibliography in Refs. 1 and 2.

[26] F. T. S. Yu, *Optics and Information Theory* (Wiley, New York, 1976); M. Ross, *Laser Receivers* (Wiley, New York, 1966); T. E. Stern, "Some Quantum Effects in Information Channels," IRE Trans. Information Theory **IT-6**, 435–440 (1960); T. E. Stern, "Information Rates in Photon Channels and Photon Amplifiers," IRE Int. Convention Record, Part 4, 182–188 (1960); J. P. Gordon, "Quantum Effects in Communication Systems," Proc. IRE **50**, 1898–1908 (1962); B. M. Oliver, "Thermal and Quantum Noise," Proc. IEEE **53**, 436–454 (1965).

[27] A. Rényi, "On an Extremal Property of the Poisson Process," Ann. Inst. Stat. Math. **16**, 129–133 (1964); J. A. McFadden, "The Entropy of a Point Process," J. Soc. Indus. Appl. Math. **13**, 988–994 (1965).

[28] Shannon and Weaver, Ref. 4, p. 19.

[29] J. C. Dainty, Proceedings of the SPSE International Conference on Image Analysis and Evaluation, July 19–23, 1976, Toronto, Canada; "The Statistics of Speckle Patterns" in *Progress in Optics*, edited by E. Wolf (North-Holland, Amsterdam, 1976), Vol. XIV.

APPENDIX B

Comments on "Spectral Estimation: An Impossibility?"

B. H. SOFFER AND RYOICHI KIKUCHI

Abstract—Spectral estimation is possible from a finite portion of the autocorrelation function. The ambiguity of a multiplicity of permissible spectra all consistent with the data is the very reason why it is possible because each possible spectrum has a different probability. Maximum entropy estimation (MEE) picks the most probable one as the estimate.

In a recent letter [1], Nitzberg questioned the possibility of making spectral estimates using only a portion of the autocorrelation function of the data because of the ambiguity arising from the multiplicity of possible spectra that are consistent with the data. Restricting the class of spectra to a particular class of network models as Nitzberg would suggest, or equivalently determining the appropriate underlying statistics (e.g., the all-pole network model in estimation theory or equivalently its underlying Gaussian statistics in complex amplitudes) [2] still would leave the ambiguity of a multiplicity of possible spectra all compatible with the given data. The multiplicity, however, is the very reason that we can employ the tools of estimation theory. Estimation theory is probabilistic and provides both an estimate and a degree of confidence in the estimate.

MEE, for example, chooses the most probable member of this set of possibilities for the estimate [2]. Entropy is defined in this context as the logarithm of the probability that a spectrum occurs. The details of this interpretation, which makes explicit the probabilistic foundation of MEE, can be found in Reference [2].

Manscript received March 23, 1979. This work was supported by the Air Force Office of Scientific Research under Contract F49620-77-C-0052.

B. H. Soffer and R. Kikuchi are with Hughes Research Laboratories, 3011 Malibu Canyon Road, Malibu, CA 90265.

0018-9219/79/1200-1672$00.75

Reply[1] *by R. Nitzberg*[2]

Soffer and Kikuchi state that spectral estimation is possible. They recommend the MEM and justify the recommendation by stating the well-known property that the technique maximizes a particular integral involving the power spectrum (entropy) under the constraint that certain other integrals involving the power spectrum are specified values (the known sampled autocorrelation values). From the viewpoint of modern statistical estimation theory, this particular estimator is one of a nondenumerable set of possible estimators. Thus is should be clear that the question of whether or not it is possible to estimate spectra cannot refer to whether *an* estimator exists but whether it can be shown that a particular one has some desirable property.

The usual estimation theory criterion for choosing a particular estimator is small error. Often, it is not known how to best select one estimator from the set of all possible; and, in order to simplify the problem, the set of estimators that is being considered is restricted. In one sense, the restriction is arbitrary but in another it is not. The restriction is made to a set with tractable mathematical properties. Thus sometimes estimators are constrained to be linear functions of the data for no reason other than nonlinear functions are hard to analyze. However, there are many well-known cases where the best linear estimators have substantially larger error than the best estimator.

This philosophy of estimation theory relates to MEM spectral estimation. It is agreed that MEM is an extremely valuable technique. Examples abound of the superiority of this technique compared to more conventional spectral estimation techniques. However, as with estimators, does the MEM procedure always have small error? As shown by Gutowski *et al.* [3], there are examples where MEM gives very bad results. A question of interest is, can it be predicted when (or if) MEM is applicable?

A main concern when using MEM should be that the concept of maximizing entropy is extremely nebulous and it may not be a reasonable criterion. Spectral estimation by MEM is a restriction to a class of estimators with tractable mathematical properties just as is the aforementioned restriction to linear estimators. This restriction is clear when the MEM procedure (as shown by Vanden Bos [4]) is recast into the all-pole network algorithm. He shows that MEM is equivalent to computing the M coefficients of an all-pole network that fits the known M values of the autocorrelation function. If the *assumption* of an M-pole network fits the problem at hand, this is a reasonable procedure. If the *assumption* is totally unreasonable it should be discarded. If it is not known whether or not the *assumption* is reasonable, one should be concerned and not adament that this is the "best" procedure. As one example of this flexibility, even when a data stream of M points is obtained, so that an M point autocorrelation function sequence can be estimated, the number of feedback coefficients of the all-pole network (the number of autocorrelation values estimated) is often taken as substantially less than M. In practice, there is not a unique MEM estimated spectrum but many depending upon how many poles are used in the estimating network. Other criteria are then imposed to choose the preferred MEM spectrum [5].

To summarize the above, there are a nondenumerable number of spectral estimators. Some of these can be phased in terms of networks with poles and zeroes. The restriction to estimation using the all-pole network (equivalent to maximizing entropy) is made primarily to simplify the mathematics. This is an extremely valuable property. The aspect of simplification is emphasized not to denigrate the technique, but to clarify under what conditions is should be expected that the procedure will work well. It is known that if, in truth, the network generating the data is an all-zero network, MEM works less well than other less easily computed procedures. If enough about the true state of nature is known so that the physics of the problem predicts troubles for MEM, do not use it. If the physics of the problem is completely unknown, spectral estimation is impossible in the sense of guaranteeing good estimation of the spectral density.

It is historically obvious that even for this latter conditon of lack of knowledge, the measured data does give considerable information about the spectrum. This is because when a truncated autocorrelation function is available, this can allow exact evaluation of some integrals of the spectrum. Note that the output a spectrum analyzer is given by integrals of this form. Specifically, when a random process is passed through a filter, the output power P_0 is given by

$$P_0 = \int_{-\infty}^{\infty} S_x(f)\, H(f)^2\, df \tag{1}$$

where $S_X(f)$ is the input process and $H(f)$ is the transfer function of the filter. By Parceval's theorem, it is also given by

$$P_0 = \int_{-\infty}^{\infty} R_X(\tau)\, R_h(\tau)\, d\tau \tag{2}$$

where $R_X(\tau)$ and $R_h(\tau)$ are the autocorrelation functions of the process and filter, respectively. If the filter impulse response is of finite duration T_F, then $R_h(\tau)$ is of duration $2T_F$. The situation stated is that $R_X(\tau)$ is known exactly for $|\tau| \leq T_1$. When $2T_F$ is less than T_1, the integral in (2) can be evaluated exactly. For a spectrum analyzer application, this is approximately equivalent to the statement that when the process is passed through a narrow-band filter tuned to any center frequency, the filter's power output can be determined as long as the reciprocal of the filter bandwidth is larger than T_1. Thus though the detailed structure of the power spectrum cannot, in general, be determined on the basis of the truncated autocorrelation function, the output of a spectrum analyzer can be computed when the spectrum analyzer's frequency resolution is not "excessive."

Further Comments by B. H. Soffer and R. Kikuchi

We agree that it would be unreasonable to use the MEM when the physics of the problem is not known. When known, the concept of the MEM is not nebulous. The MEM is not limited to the form introduced by Burg based on Gaussian statistics; other forms of entropy may be dictated by the problem at hand [2].

References

[1] R. Nitzberg, "Spectral estimation: An impossibility?" *Proc. IEEE*, vol. 67, pp. 437-438, Mar. 1979.

[2] R. Kikuchi and B. H. Soffer, "Maximum entropy image restoration I. The entropy expression," *J. Opt. Soc. Amer.*, vol. 67, pp. 1656-1665, Dec. 1977.

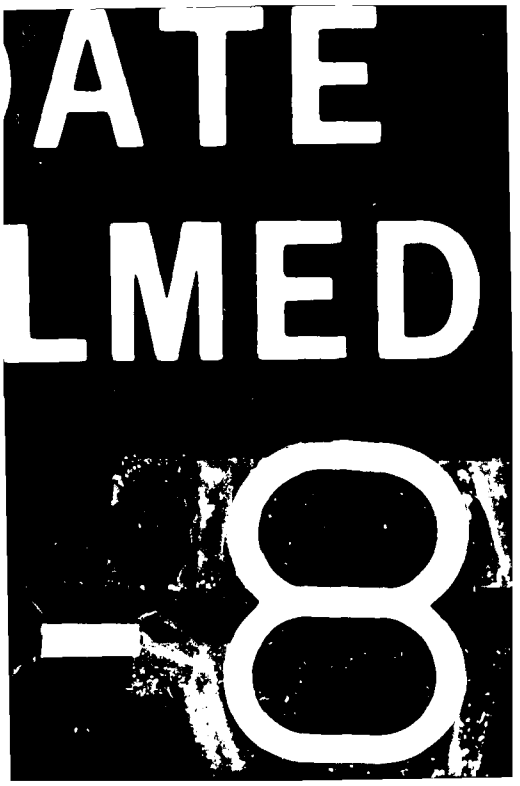
[3] R. R. Gutowski, E. A. Robinson, and S. Treitel, "Spectral estimation: Fact or fiction," *IEEE Trans. Geosci. Electron.*, vol. GE-16, pp. 80-84, Apr. 1978.

[4] A. Vanden Bos, "Alternate interpretation of maximum entropy spectral estimation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 493-494, July 1971.

[5] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Math.*, vol. 21, pp. 243-247, 1969.

[1] Manuscript received June 25, 1979.

[2] R. Nitzberg is with General Electric Company, Syracuse, NY 13221.


ATE
LMED
8